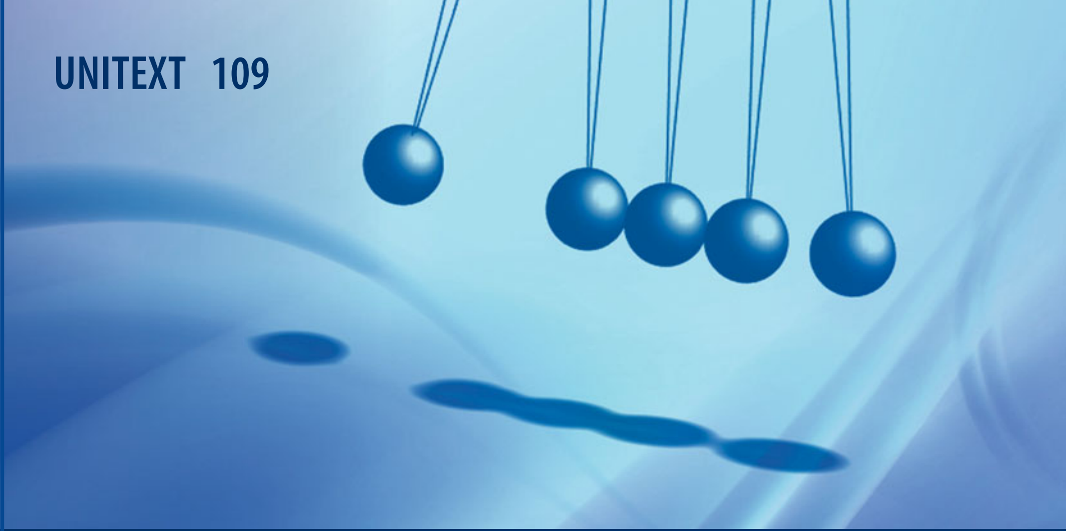


UNITEXT 109

Andreas Knauf

Mathematical Physics: Classical Mechanics

Springer

UNITEXT - La Matematica per il 3+2

Volume 109

More information about this series at http://www.springer.com/series/5418

Andreas Knauf

Mathematical Physics: Classical Mechanics

Springer

Andreas Knauf
Department of Mathematics
Friedrich-Alexander University
Erlangen-Nürnberg
Erlangen
Germany

Translated by Jochen Denzler, Department of Mathematics, University of Tennessee, Knoxville, Tennessee, USA

ISSN 2038-5714 ISSN 2532-3318 (electronic)
UNITEXT - La Matematica per il 3+2
ISSN 2038-5722 ISSN 2038-5757 (electronic)
ISBN 978-3-662-55772-3 ISBN 978-3-662-55774-7 (eBook)
https://doi.org/10.1007/978-3-662-55774-7

Library of Congress Control Number: 2017952919

Mathematics Subject Classification (2010): 70-01, 37-01, 37J05, 37J40

Translation from the German language edition: *Mathematische Physik: Klassische Mechanik* by Andreas Knauf, Springer-Lehrbuch Masterclass, © Springer, 2nd edition 2017. All rights Reserved.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH, DE
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Contents

Remarks on Mathematical Physics

Themes and Goals

> "*The laws of nature are constructed in such a way as to make the universe as interesting as possible.*" FREEMAN DYSON, in *Imagined Worlds* (1997)

In mathematical physics, one attempts, beginning with the fundamental equations and assumptions of physics (like Newton's equation, the Boltzmann distribution, or the Schrödinger equation), to derive facts of physics by means of mathematics.

So it is the problem of physics that is center stage (for instance, the question about the stability of the solar system, the reason for the existence of crystals, or the localization of electrons in an amorphous solid).

The methods needed to solve the respective problems can, in their majority, be classified as analysis or geometry, but algebraic techniques also play a role. Roughly, the mathematical correspondence

- to classical mechanics is the theory of ordinary differential equations,
- to quantum mechanics is functional analysis, and
- to (classical) statistical mechanics is probability theory.

However, a series on theoretical physics should also include electrodynamics and thus, mathematically speaking, the theory of Maxwell's equations, which are linear partial differential equations. The general theory of relativity, one of the foundations of modern physics, leads, like many other questions, to a nonlinear partial differential equation. Quantum field theory relies on a variety of analytic, geometric, as well as algebraic methods.

Given the vast extent of the area, the question arises how one could possibly gain some ground within a reasonable time and whether the study of mathematical physics is worthwhile.

The present first volume of a projected three-volume course on mathematical physics offers an answer to the first question.

As for the second question, everybody needs to decide for themselves. Students of mathematics and physics often have different motives in this respect:

- In the regular sequence of lectures on *theoretical physics*, a mathematically rigorous foundation cannot be provided for reasons of time. Thus, for instance, Schrödinger operators are treated like finite matrices as a matter of necessity. Here, mathematical physics can serve as a reasonable complement.
  The price to be paid for a higher mathematical rigor is that a course on mathematical physics cannot discuss as large as variety of physical phenomena within the time decreed for a bachelor's degree as could be done in a course on theoretical physics. Instead, we will focus on conceptual foundations and investigate representative models.
- In the study of *mathematics*, a deductive development of mathematical notions is chosen for good reason.
  In this context, mathematical physics, which is oriented at problems, not at methods, can motivate the practical relevance of these notions—for instance, the dynamical relevance of the spectral decomposition of a self-adjoint operator.

Another reason for the interest in mathematical physics is a phenomenon that Eugene Wigner made into the title of an essay of 1960, namely

"*The unreasonable effectiveness of mathematics in the natural sciences*"

Or, quoting Albert Einstein:

"*How is it possible that mathematics, which is an independent product of human thought, independent of all practical experience, is such an excellent fit for objects in reality?*"[1]

The immediate application of the quote by Einstein is to "all objects in reality". However, it is in physics that it is confirmed best. Here, mathematical structures (like differential geometry for the theory of relativity or group theory for quantum field theory) lead the way to an appropriate theory, often prior to observations in experiments.[2]

In contrast, for instance in biology, which is considered the new lead science, structures that have arisen in natural history are key. Even if many phenomena can be modeled mathematically, the predictive value of the models is limited. The same applies even more for instance for economic sciences.

The diversity of technological inventions can also be described in the language of mathematics, and here, too, one works more by problems than by methods; the distinction is, however, in the goal. The primary goal of mathematical physics is the insight into phenomena of nature, whereas in "Mathematics for Technology", the ultimate purpose is the simulation and optimization of structures and processes.

---

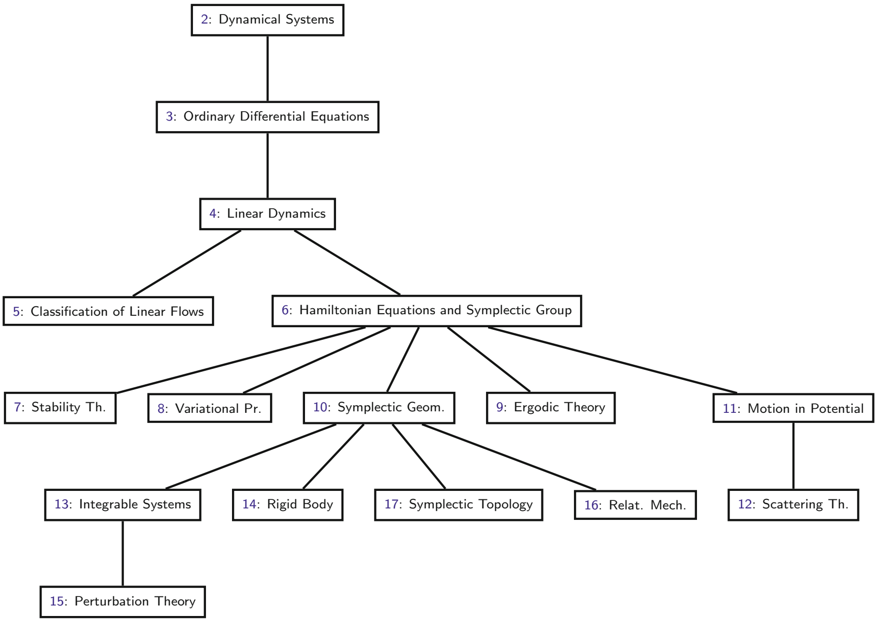
[1] A. Einstein: Geometrie und Erfahrung. Festvortrag, gehalten an der Preussischen Akademie der Wissenschaften zu Berlin, am 27. Januar 1921. Berlin: Julius Springer 1921.

[2] Conversely, the book [Lev] by MARK LEVI shows in an entertaining manner that "proofs by physics" can be given for facts of mathematics.

Contents of the Book "Classical Mechanics"

> *"In the prescientific era, the Greek term μηχανη′ (mêchanê) refers to a construction, artifice, or an (illegitimate) trick. If treaties in ancient Greece wanted to outlaw deceitful conduct, they would prohibit the use of τε′χνη (technê) or mêchanê, of deceitfulness and malice." [Me]*, page 129

This is the first in a textbook series on mathematical physics that is projected to have three volumes. The following tree gives the first approximation to the logical interdependence among the chapters of the book.



By example, the following selection of material could be the basis of a four-hour course (provided basic knowledge about ordinary differential equations can be assumed):

- Chapter 2: Dynamical Systems
- Chapter 6: Hamiltonian Equations and Symplectic Group
- Chapter 8: Variational Principles
- Chapter 9: Ergodic Theory
- Chapter 10: Symplectic Geometry
- Chapter 13: Integrable Systems

Contents of the Series

This projected series of textbooks is intended as a basis for a three-semester course (4 hours each) on mathematical physics, like they are offered at several universities in German-speaking countries. The material is limited in a way that can be integrated into a curriculum in either mathematics or physics.

Moreover, the three volumes are largely independent of each other, which allows to integrate them more easily in different curricula.

Explicitly assumed prerequisites are only the lectures on Analysis and Linear Algebra. The main prerequisites from, e.g., Differential Geometry, Group Theory, Topology, and Probability are gathered in appendices. *Essentials* from, say, the theory of ordinary differential equations or functional analysis will be introduced where needed. Students who have these prerequisites may skip the corresponding chapters.

The volumes are appropriate for individual study.

This projected series cannot replace the well-known four volume treatises, *Methods of Modern Mathematical Physics* by M. Reed and B. Simon, or the books on *Mathematical Physics* by W. Thirring. Whereas the former pioneered the mathematization of quantum mechanics, the latter went the entire length from the basic equations to the proof of properties of physically realistic models.

Here, we start at a slightly lower level of mathematical sophistication, but at the same time attempt to reflect the breadth of present questions and to lay the foundations for an understanding of more specialized literature.

Chapters marked by an asterisk (*) are mathematically more advanced, but will not be assumed in the sequel.

The exercises (of which this volume contains over 100) will to some part be complemented with hints, as their level of difficulty varies considerably. Solutions are gathered in an appendix. The illustrations (some 340 for the present volume) are, to the extent possible, quantitatively exact.

Acknowledgement

The projected series of textbooks originates in the lecture series “Mathematical Physics” that was established by Ruedi Seiler in the Dept of Mathematics of the Technische Universität (University of Technology) Berlin. I myself, as well as the series of textbooks, owe him a lot.

Robert Schrader (1939-2015)[3], who directed my Master Thesis (Diplomarbeit) and Ph.D. Thesis at the Department of Physics at the Freie Universität (Free

[3] The International Association of Mathematical Physics (IAMP, www.iamp.org) published an obituary for Robert Schrader in its News Bulletin of April 2016.

University) Berlin, has shaped my view of mathematical physics decisively. He was an excellent scientist. For me, he was a mentor, colleague, and friend. I dedicate this volume to his memory.

I am grateful to Ms. Irmgard Moch, who did the detective work of deciphering my handwriting and typed the German edition of this book. Christoph Schumacher contributed several exercises, among other things. Viviane Baladi, Tanja Dierkes, Jacques Féjoz, Daniel Matthes, Herbert Lange, Johannes Singer, Zhiyi Tang, Stefan Teufel, Stephan Weis, and numerous other colleagues discovered typos or mistakes in the manuscript, or contributed otherwise to improving it.

I am grateful to Ms. Allewelt, Ms. Herrmann, and Mr. Heine from Springer Publisher for their friendly help in publishing this book.

Concerning this English edition, I am very grateful to Jochen Denzler (Department of Mathematics, University of Tennessee) for having done a superb job in translating the German edition. He was also extremely helpful in improving the text.

Any remaining mistakes are of course my responsibility. I appreciate being notified of such mistakes.

Erlangen, Germany Andreas Knauf
July 2017

Notation

Subsets: If A and B are sets, A is called a *subset* of B (written as $A \subseteq B$) if $x \in A \Rightarrow x \in B$. In particular, it is true that $B \subseteq B$. *Strict inclusion* $A \subsetneq B$ means $A \subseteq B$, but $A \neq B$. In mathematical literature, one also finds the symbol $A \subset B$ for subsets.

Power Sets: If A is a set, then the *power set of* A is

$$2^A := \{B | B \subseteq A\}.$$

Synonymous notations in the literature are $\mathfrak{P}(A)$ and $\mathcal{P}(A)$.

Functions: For $f : M \to N$ and $A \subseteq M$, we define $f(A) := \{f(a) | a \in A\}$.
For $B \subseteq N$, we define $f^{-1}(B) := \{m \in M | f(m) \in B\}$. For $b \in N$, we will define $f^{-1}(b) := f^{-1}(\{b\})$.

Numbers: $\mathbb{N} = \{1, 2, \ldots\}$ is the set of positive integers, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$,
$\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$ is the ring of integers.
$\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are the fields of rational, real, and complex numbers, respectively.
For a field $\mathbb{K}$, the symbol $\mathbb{K}^*$ will denote the multiplicative group, $\mathbb{K}^* := \mathbb{K} \backslash \{0\}$. Finally,

$$\mathbb{R}^+ := \{x \in \mathbb{R} | x > 0\} = (0, \infty).$$

We write $\imath$ rather than i for the imaginary unit.

Intervals: $I \subseteq \mathbb{R}$ is an *interval* if, for $y < z \in I$, every point $x \in \mathbb{R}$ with $y \le x \le z$ belongs to I. For $a \in \mathbb{R} \cup \{-\infty\}$ and $b \in \mathbb{R} \cup \{+\infty\}$, we denote

$$(a,b) := \{x \in \mathbb{R} | x > a,\ x < b\},\ (a,b] := \{x \in \mathbb{R} | x > a,\ x \le b\},$$

etc.

(Synonymous notations are $]a,b[= (a,b)$, $]a,b] = (a,b]$, etc.).

We occasionally also use (generalized) intervals like $[0, +\infty] = [0, \infty) \cup \{+\infty\}$.

Matrices: $Mat(m \times n, \mathbb{K})$ denotes the $\mathbb{K}$-vector space of $m \times n$ matrices with entries from the field $\mathbb{K}$, and $Mat(n, \mathbb{K})$ denotes the ring $Mat(n \times n, \mathbb{K})$.

Spheres and Balls: For $d \in \mathbb{N}_0$, we define $S^d := \{x \in \mathbb{R}^{d+1} | \parallel x \parallel = 1\} = \partial B^{d+1}$, i.e., the boundary of the closed unit ball; here $B^d_r := \{x \in \mathbb{R}^d | \parallel x \parallel \le r\}$ is the closed unit ball of radius $r > 0$, and $B^d := B^d_1$.

We write $S^1 \subset \mathbb{C}$ for $\{c \in \mathbb{C} | |c| = 1\}$, but also $S^1 := \mathbb{R}/\mathbb{Z}$ (with the identification $[x] \mapsto exp(2\pi\imath x)$) for the multiplicative and additive group, respectively.

The Greek Alphabet

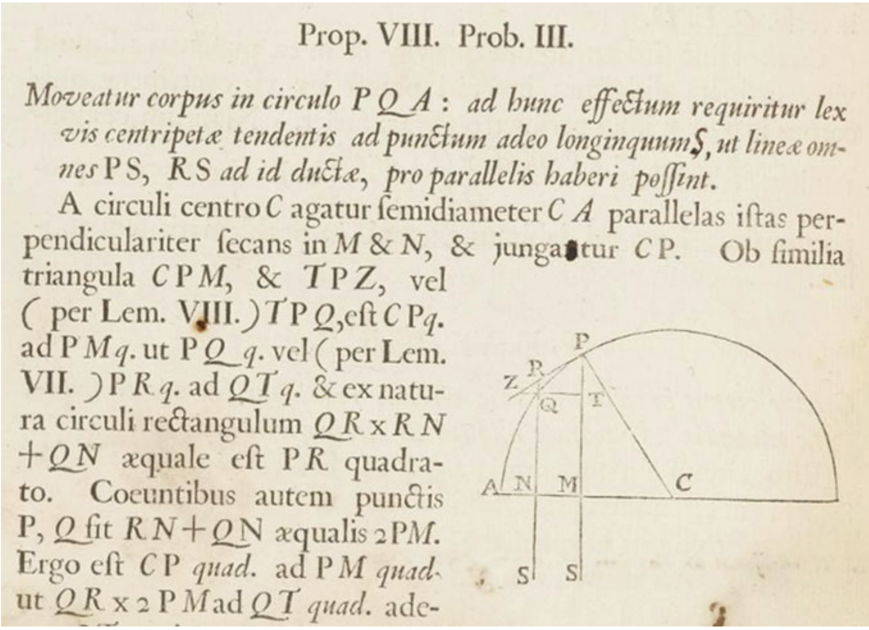
a) Lower case letters

α	Alpha	ζ	Zeta	λ	Lambda	π	Pi	φ, ϕ	Phi
β	Beta	η	Eta	μ	My	ρ, ϱ	Rho	χ	Chi
γ	Gamma	θ, ϑ	Theta	ν	Ny	σ, ς	Sigma	ψ	Psi
δ	Delta	ι	Jota	ξ	Xi	τ	Tau	ω	Omega
ϵ, ε	Epsilon	κ	Kappa	o	Omikron	υ	Ypsilon		

b) Capital letters (to the extent that they are distinct from capitals in the Latin alphabet)

Γ	Gamma	Θ	Theta	Ξ	Xi	Σ	Sigma	Φ	Phi	Ω	Omega
Δ	Delta	Λ	Lambda	Π	Pi	Υ	Ypsilon	Ψ	Psi		

Chapter 1
Introduction



Prop. VIII. Prob. III.

Moveatur corpus in circulo PQA: ad hunc effectum requiritur lex vis centripetæ tendentis ad punctum adeo longinquum S, ut lineæ omnes PS, RS ad id ductæ, pro parallelis haberi possint.

A circuli centro C agatur ſemidiameter CA parallelas iſtas perpendiculariter ſecans in M & N, & jungatur CP. Ob ſimilia triangula CPM, & TPZ, vel (per Lem. VIII.) TPQ, eſt CPq. ad PMq. ut PQq. vel (per Lem. VII.) PRq. ad QTq. & ex natura circuli rectangulum QR x RN + QN æquale eſt PR quadrato. Coeuntibus autem punctis P, Q fit RN + QN æqualis 2PM. Ergo eſt CP quad. ad PM quad. ut QR x 2PM ad QT quad. ade-

From Newton's own copy of the first edition of his book *Philosophiæ Naturalis Principia Mathematica*. Reproduced by kind permission of the Syndics of Cambridge University Library

How it all Got Started

> And he was told to tell the truth, otherwise one would have recourse to torture. [He replied:] I am here to obey, but I have not held this opinion after the determination was made, as I said.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_1

Judicial protocol of the Holy Office of Inquisition on the GALILEI case (1633)[1]

A bit more than 50 years lapsed between the guilty verdict on Galilei and the publication of Newton's *Principia*. During this time, modern science was established, with classical mechanics as the leading science. We will begin this introduction with the solution of the equation of motion for the planets, which amounts to confirming Galileo Galilei's heliocentric cosmology and making it more precise.

Both the research of differential equations and the physics-based celestial mechanics go back to Isaac Newton (1643–1727).

- Isaac Newton the *mathematician* is known (together with Leibniz) as the founder of differential calculus.
- Isaac Newton the *physicist* lent his name to the law

$$\text{Force} = \text{Mass} \times \text{Acceleration}. \tag{1.1}$$

Both of these activities of Newton's are interconnected. Namely, if

- $x(t) \in \mathbb{R}^3$ is the *position* of a particle at time $t \in \mathbb{R}$, then (using Newton's notation for time derivatives) $\dot{x}(t) = \frac{dx}{dt}(t) \in \mathbb{R}^3$ is its *velocity* and $\ddot{x}(t) = \frac{d^2x}{dt^2}(t) \in \mathbb{R}^3$ is its *acceleration* at that moment of time.
- On the other hand, the *force* F can depend on the position and velocity of the particle, and also directly on the time, so that equation (1.1) has the form

$$\boxed{F(x, \dot{x}, t) = m\ddot{x}}.$$

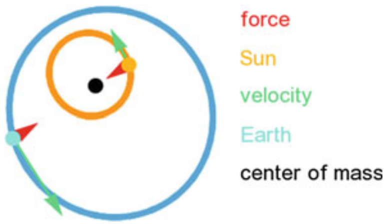
It is assumed here that the force function F is known.

This is an example of a differential equation, as it is an equation satisfied by the unknown (vector-valued) function $t \mapsto x(t)$ and its derivatives.

For instance, on the earth (with its center[2] located at x), with mass $m > 0$, the force is

$$F(x) = -m\gamma \frac{x}{\|x\|^3} \qquad \left(x \in \mathbb{R}^3 \setminus \{0\}\right), \tag{1.2}$$

where we have used the Euclidean norm $\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$, and where we have assumed for simplicity that the sun is at rest in the origin of the coordinate system. A more detailed study shows that the true 2-body problem, in which earth and sun move around their common center of mass, can be



[1] See SOBEL [Sob2, Section 24].

[2] In Problem 12.37 on page 308, we show that for the centrally symmetric distribution of mass, the gravitation has the same effect as if all mass were concentrated in the center.

reduced to the central force problem just discussed if in (1.2), the mass m of the earth is replaced with the *reduced mass* $\frac{mM}{m+M} < m$.[3] The positive constant γ is the product of the gravitational constant and the mass of the sun. So after cancelling m, equation (1.1) has the form

$$\boxed{\ddot{x} = -\gamma \frac{x}{\|x\|^3}}. \tag{1.3}$$

Newton solved this differential equation, and in doing so, he derived Kepler's laws of planetary motion, which had previously only been abstracted from observational data, as a consequence of the foundational mechanical laws (1.1) and (1.2). This was the first triumph of the new natural science—published in his main opus "Philosophiæ Naturalis Principia Mathematica" (shortly: *Principia*) [Ne] in 1687.

Newton was well aware of the significance of his insight; and as he also leaned towards the mysterious, next to mathematics and physics, he encoded the Latin sentence *data aequatione quotcunque fluentes quantitates involvente, fluxiones invenire et vice versa* in an anagram[4]

`6accdae13eff7i319n4o4qrr4s8t12ux`.

A free translation of that sentence reads

> It is useful to solve differential equations.

Derivation of Kepler's Laws

Let us follow Newton's advice and solve (1.3).

1. First we observe that the planet will remain, for all times, within the *orbital plane* spanned by its initial position and its initial velocity[5]; this is because

$$\frac{\mathrm{d}}{\mathrm{d}t}[x \times \dot{x}] = \dot{x} \times \dot{x} + x \times \ddot{x} = \dot{x} \times \dot{x} - \gamma \frac{x \times x}{\|x\|^3} = 0, \tag{1.4}$$

   so the vector $x(t) \times \dot{x}(t) \in \mathbb{R}^3$, which is orthogonal to this plane, is constant in time.
2. Now it is useful to describe the position $x(t)$ in this orbital plane by a complex number $z(t)$, where $z(t) = x_1(t) + \mathrm{i}x_2(t)$, provided $x \times \dot{x}$ points in the third coordinate direction (which we may assume with no loss of generality).
   Then in *polar coordinates*, one has $z(t) = r(t)e^{\mathrm{i}\varphi(t)}$, and therefore

$$\dot{z} = (\dot{r} + \mathrm{i}\, r\dot{\varphi})e^{\mathrm{i}\varphi} \text{ and } \ddot{z} = \big(\ddot{r} - r\dot{\varphi}^2 + \mathrm{i}(2\dot{r}\dot{\varphi} + r\ddot{\varphi})\big)e^{\mathrm{i}\varphi}.$$

---

[3] See Example 12.39 on page 310.

[4] See http://www.newtonproject.sussex.ac.uk/view/texts/normalized/NATP00180. Quotation from the introduction of ARNOL'D's book [Ar3]. Interesting biographical notes on Newton can be found in [Ar6].

[5] If $\dot{x}$ is parallel to x, we obtain a straight line instead of a plane.

After division by $e^{i\varphi}$ and separation of real and imaginary parts, Newton's force equation, $\ddot{z} = -\gamma \frac{z}{|z|^3}$, leads therefore to the coupled real differential equations

$$\text{(I): } \ddot{r} - r\dot{\varphi}^2 + \frac{\gamma}{r^2} = 0, \ \text{(II): } 2\dot{r}\dot{\varphi} + r\ddot{\varphi} = 0. \tag{1.5}$$

3. Because $\frac{d}{dt}(r^2\dot{\varphi}) = r(2\dot{r}\dot{\varphi} + r\ddot{\varphi}) = 0$, the quantity $\ell := r^2\dot{\varphi} = \text{const}$ is a constant of motion. By definition, the product of this quantity with the mass m of the planet is the *angular momentum.* So this quantity is constant in time.
4. Substituting $\dot{\varphi} = \ell/r^2$ into (I) produces the equation

$$\ddot{r} - \frac{\ell^2}{r^3} + \frac{\gamma}{r^2} = 0.$$

Here, too, we can find a constant of motion; for if we let

$$U_\ell(r) := \frac{\ell^2}{2r^2} - \frac{\gamma}{r} \text{ and } H_\ell(r, \dot{r}) := \tfrac{1}{2}\dot{r}^2 + U_\ell(r),$$

we have $\frac{d}{dt} H_\ell\big(r(t), \dot{r}(t)\big) = \dot{r}\left(\ddot{r} - \frac{\ell^2}{r^3} + \frac{\gamma}{r^2}\right) = 0$, hence H_ℓ is constant in time: $H_\ell\big(r(t), \dot{r}(t)\big) = H_\ell\big(r(0), \dot{r}(0)\big) =: E$. In physics, the product $E\,m$ is interpreted as the *total energy* of the planet, and all real numbers can occur as values of the energy.
5. Now initially, we are less interested in solving the differential equation

$$\dot{r} = \pm\sqrt{2(E - U_\ell(r))}, \tag{1.6}$$

namely how the radius depends on time, than we are in the shape of the orbit $R(\varphi) := r(t(\varphi))$. If we assume that $\ell = r^2\dot{\varphi} \neq 0$, we can switch to φ as the independent variable. Then, from (1.6), we obtain

$$\frac{dR}{d\varphi} = \frac{\dot{r}}{\dot{\varphi}} = \frac{\pm R^2\sqrt{2(E - U_\ell(R))}}{\ell},$$

hence by separating variables and substituting U_ℓ,

$$\int \frac{\ell\, dR}{\pm R\sqrt{2ER^2 + 2\gamma R - \ell^2}} = \int d\varphi' = \varphi - \varphi_0.$$

Using the constants $e := \sqrt{1 + \frac{2E\ell^2}{\gamma^2}}$ and $p := \ell^2/\gamma$, the integrand on the left hand side can be transformed into

$$\frac{\ell}{R\sqrt{2ER^2 + 2\gamma R - \ell^2}} = \frac{p/R}{\sqrt{e^2R^2 - (p - R)^2}}.$$

Using a table of integrals, we conclude $\arccos\left(\frac{p/R(\varphi) - 1}{e}\right) = \varphi - \varphi_0$, or

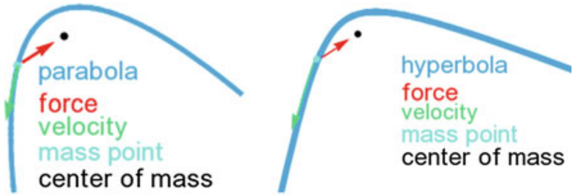
$$R(\varphi) = \frac{p}{1 + e\cos(\varphi - \varphi_0)}. \tag{1.7}$$

But this is the equation of a conic, in which the constant e is called the *eccentricity* and p the *parameter* of the conic. The angle $\varphi - \varphi_0$ is called *true anomaly* in the astronomy literature.

Results

1. We obtain, for angular momentum $\ell \neq 0$, depending on the energy E, the following:
   - For $E < 0: \;\; 0 \leq e < 1$ (an ellipse), with the circle as a special case for $e = 0$.
   - For $E = 0: \;\; e = 1$ (a parabola)
   - For $E > 0: \;\; e > 1$ (a hyperbola)

   We have thus derived **Kepler's first law**.



2. **Kepler's second law** states that the segment connecting sun and planet will swipe out the same area during the same time.
   This law follows from the fact that $\ell = r^2\dot{\varphi}$ is constant, because the area swiped out during the time interval $[t_1, t_2]$ is

$$\int_{\varphi_1}^{\varphi_2} \tfrac{1}{2} R(\varphi)^2 \, \mathrm{d}\varphi = \tfrac{1}{2} \int_{t_1}^{t_2} r(t)^2 \frac{\mathrm{d}\varphi}{\mathrm{d}t}(t) \, \mathrm{d}t = \tfrac{1}{2}\ell \, (t_2 - t_1).$$

3. Similarly, we confirm **Kepler's third law**, according to which the squares of the orbital periods[6] (of different planets) are proportional to the third powers of the major semiaxes.
   For a solution $t \mapsto x(t)$ of the force equation (1.3), we substitute $X(t) := c_x x(t/c_t)$ with positive parameters c_x and c_t. It is exactly when $c_t^2 = c_x^3$ that the curve $t \mapsto X(t)$ again satisfies the force equation; indeed

$$\frac{\mathrm{d}^2 X}{\mathrm{d}t^2}(t) = \frac{c_x}{c_t^2} \frac{\mathrm{d}^2 x}{\mathrm{d}t^2}(t) \text{ and } \frac{X}{\|X\|^3} = c_x^{-2} \frac{x}{\|x\|^3}.$$

   This proves the law of scaled orbits of equal shape.

---

[6] To memorize which quantity to square: *Times Square*.

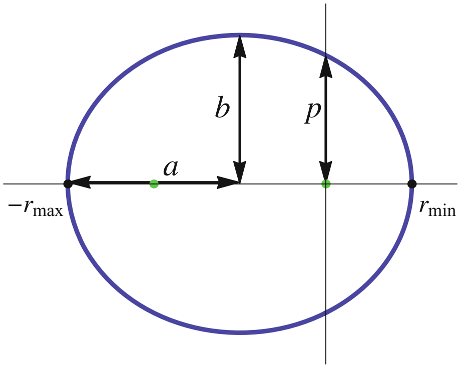
1.1 Exercise (**Kepler's Third Law**) Show that the orbital period on a Kepler ellipse depends only on its major semiaxis.[7]

Hint: Show first that

(a) for the minimal and maximal distances $r_{\min} := R(\varphi_0)$ and $r_{\max} := R(\varphi_0 + \pi)$, the length of the major semiaxis equals the arithmetic mean $a := \frac{1}{2}(r_{\min} + r_{\max})$, whereas the minor semiaxis equals the geometric mean $b := (r_{\min}\, r_{\max})^{1/2}$.

(b) Due to Kepler's second law, the orbital period is $2\pi ab/\ell$.

(c) Next use the relation $\ell^2 = \gamma^2 \frac{e^2-1}{2E}$ between angular momentum and eccentricity. ◊



So we have solved the differential equation of the Kepler problem.[8]

While this may not have become transparent, we have profited from the *symmetry* of the system under consideration, namely the invariance of (1.3) under orthogonal transformations of $\mathbb{R}^3$, by noticing that certain quantities were constant in time; and this allowed us to reduce the number of variables. Newton's equation with a central force is always integrable.

In (1.2), at most second derivatives of x with respect to time occur, and accordingly, we need to prescribe, as *initial conditions* at time 0, the initial position $x(0) \in \mathbb{R}^3$ and the initial velocity $\dot{x}(0) \in \mathbb{R}^3$. Then there *exists for all times a unique solution of* (1.3) (except in the case of collision with the central mass at $x = 0$, which occurs when $x(0)$ and $\dot{x}(0)$ are linearly dependent).

The *n*-Body Problem

This success of Newton's is contrasted with a failure.

One may study, more generally, the motion of n mass points that attract each other. For instance, consider the solar system consisting of the sun and several planets.

Since the force on the ith body (with position $q_i \in \mathbb{R}^3$) equals the product of its mass m_i and its acceleration $\ddot{q}_i$, one obtains[9] from Equations (1.1) and (1.2) the system of differential equations

$$\ddot{q}_i = \sum_{j \neq i} m_j \frac{q_j - q_i}{\|q_j - q_i\|^3} \qquad (i = 1, \ldots, n). \tag{1.8}$$

---

[7] It follows from the formulas for the semiaxes that their ratio b/a is $1 + \mathcal{O}(e^2)$, whereas the ratio $r_{\min}/r_{\max}$ depends on the eccentricity e in the form $\frac{1-e}{1+e} = 1 - 2e + \mathcal{O}(e^2)$. Planetary orbits with eccentricities $e < \frac{1}{4}$ are therefore, in first approximation, circles, where the sun is shifted by order e from the center. The ellipse in the figure has eccentricity $e = \frac{1}{2}$, whereas the earth has $e = 0.017$.

[8] The parametrization of the conics by time is obtained by elliptic integrals.

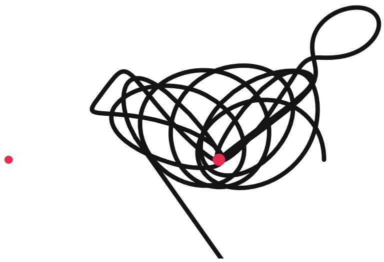
[9] in units that make the constant of gravity 1.

It is the fact that the planets are of so much smaller mass compared to the sun that results in the approximate solution from the two-body problem being so precise.

But nevertheless, Jupiter, the heaviest planet, has almost 1/1000 the mass of the sun. So one can expect that already after about 1000 years, Jupiter's presence has a notable effect on the orbit of the earth. So the solution of the next simplest problem of three bodies would be of special interest to us. However, the 3-body problem resisted all attempts to solve it for 200 years after the publication of 'Principia'.[10]

In 1885, the French mathematician Henri Poincaré was awarded a prize that had been offered by the Swedish King Oscar II for the solution of this problem. However, Poincaré had not solved the problem in its generality, but rather had found indications that attempts at a series solution would have to be divergent (DIACU and HOLMES [DH] tell this interesting story.)

1.2 Remark (**Three-Body Problem**)
As a matter of fact, the orbits of the *restricted 3-body problem* are already very complicated. In the restricted 3-body problem, a satellite with negligibly small mass is moving in the gravitational field of two bodies that circle about their common center of mass.



In the numerical solution of the differential equation depicted on the right, one can see the shape of the orbit of the satellite, as seen in a coordinate system that rotates along with the massive bodies, whose position is therefore fixed in that coordinate system. ◊

This observation by Poincaré gave rise to a new era in which more emphasis is placed on qualitative properties of differential equations. For instance, one may ask whether the solar system is stable or not, or whether it is possible to capture celestial bodies in the absence of friction forces, etc. In this book, both types of questions, those about explicit solutions to differential equations, and those about their qualitative properties, will find a place.

Constraints and Friction

It is the goal of classical mechanics to conclude from the kind of forces acting between the massive bodies to the kind of motion they undergo. A first task in this process consists of setting up the equation of motion.

Let us consider the motions of a ball. It consists of a huge number of atoms. However, inasmuch as their distances are constant in time, the ball can be considered as a rigid body, and six equations of second order suffice to describe its motion in space. Three of them describe the position of its center of mass, and three describe the rotation with respect to its initial configuration.[11]

---

[10] See Chapter 11, beginning on page 256.

[11] See Chapter 14, beginning on page 365.

If the ball rolls on a surface, this constraint reduces the number of degrees of freedom by one. If it rolls without sliding, then a motion of the center of mass is only possible if the ball rotates at the same time. In this case, the question arises whether the ball can reach any prescribed point in its configuration space by appropriate motions.

Whereas in the case of the motion of the earth around the sun, we can neglect friction in good approximation, this is not the case for a ball rolling on the earth. Here the deceleration is proportional to the speed. In the course of time, the ball dissipates energy into the environment and thus gradually comes to rest.

We will predominantly discuss Hamiltonian systems, in which the energy is a quantity that is conserved in time. In doing so, we will see that Hamiltonian systems allow a geometric description; this geometry of the phase space, called symplectic geometry, differs significantly from the more customary Riemannian geometry.

Nevertheless, in particular at the beginning, we also want to consider more general dynamical systems; for instance those that are losing energy. It is only in this way that we can clearly recognize the special nature of Hamiltonian dynamics.

Discrete Dynamical Systems

Here is one last example: In high energy physics, particles are accelerated to high energies in order to let them collide afterwards. But for various reasons, often one does not want to make particles collide immediately. This is why storage rings are being used, in which the particles that had been accelerated to almost the speed of light are forced onto a circular orbit by means of a magnetic field.

A constant magnetic field produces a spiral-shaped orbit in which the component of the velocity parallel to the magnetic field is constant.[12] After few orbits, the charged particles would therefore leave the storage ring. To prevent this, more complicated configurations of magnets are used.

In order to study these configurations, one can measure location and velocity of the particle perpendicular to the direction of the ray at one particular position of the ring. This way we obtain a mapping that allows to obtain, from position q_n and velocity v_n of the particle in the n^{th} orbit, the values of these same quantities in the $(n+1)^{\text{th}}$ orbit:

$$q_{n+1} = f_1(q_n, v_n),\ v_{n+1} = f_2(q_n, v_n),$$

or briefly,

$$x_{n+1} = F(x_n) \text{ with } x_k := \binom{q_k}{v_k} \text{ and } F := \binom{f_1}{f_2}.$$

In order to determine the mapping F, one has to calculate the effects of the magnets on the relativistic particles.

Where are the particles by the end of the time of storage, for instance after 10^8 orbits? To answer this question, we must, in principle, iterate the above mapping. We set

$$F^{n+1} := F \circ F^n \text{ with } F^0 := \mathrm{Id}.$$

[12]This will be shown in Section 6.3.3 on page 122.

If F is invertible, then we define analogously $F^{-n-1} := F^{-1} \circ F^{-n}$.

This way, we obtain kind of a stroboscopic model of the motion. In this model, the time parameter n takes on values in the integers $\mathbb{Z}$, whereas in the previous case, we had $t \in \mathbb{R}$.

Again one could have conservation of energy in good approximation, or else energy could be lost, for example by radiation.

One engineering task consists of designing the magnets in such a way that one obtains stability, such that the particles do not leave the storage ring. In doing so, it would be calculationally ineffective to actually do 10^8 iterations of F. After all, we have to show stability for a continuum of initial conditions. A deeper analysis of dynamics is needed.

Let us assume that $F(0) = 0$, in other words, those particles that start without deviation of location and velocity will also return that way. A first approximation then consists of the linearization of F, i.e., one needs to investigate the matrix $M := \mathrm{D}F(0)$. If this matrix were to have only complex eigenvalues whose absolute values are less than 1, then the motion would be stable.

However, in the case of conservation of energy, this will not happen. The best that we can achieve constructively here is marginal stability, in which all complex eigenvalues are located on the unit circle. In this case, can we still conclude from M to F? This is indeed possible with Hamiltonian perturbation theory, according to the famous theorem of Kolmogorov, Arnol'd, and Moser[13] which even applies for infinite time. Indeed, as one varies the parameters of F, i.e., the configuration of the magnets, one discovers that for particular values of the parameters, a bifurcation from unstable to stable behavior occurs. So, in principle, the engineering task is solvable.

Additionally various interactions between particles within the beam threaten stability, in particular for large intensities. Then their collective motion is modeled by the Vlasov equation, a Hamiltonian partial differential equation.

Figure 1.1 shows how instabilities appear for large charges.

Relation to Statistical Mechanics and Quantum Mechanics

Whereas the rigid body, which consists of many atoms, could be described by merely six coupled differential equations, this is not the case for gases. Nevertheless, gases can be described very effectively in statistical mechanics. However, this is no longer done on a deterministic basis, but on a probabilistic one. One assumes that the equilibrium state of the gas can be described by few parameters like density and energy; it is a probability measure that is evenly distributed on the manifold defined by these parameters. This successful assumption, however, needs to be justified by analyzing the ergodic properties of the mechanical system.

In reality, nature is not classical, but quantized. The Newtonian equation of classical mechanics only appears in an asymptotics of the quantum mechanical Schrödinger

[13] KAM theory will be discussed in Section 15.4, beginning on page 412.

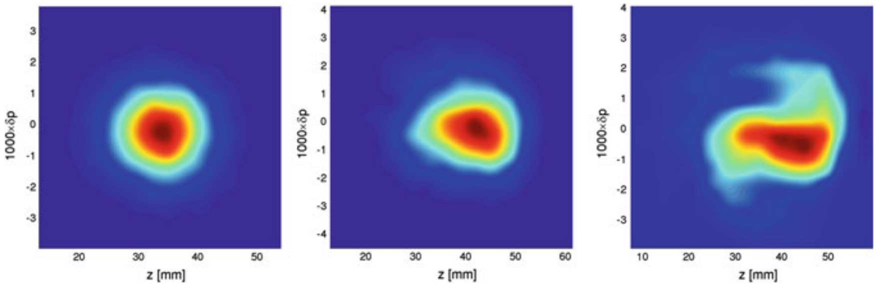


Figure 1.1 Beam instability. Left: charge 1 picocoulomb, middle: charge 8 pC, right: charge 12 pC. Images: courtesy of Andy Wolski (Liverpool)

equation. Nevertheless, nobody would have the idea, for instance, to study the mechanical properties of a bicycle by means of quantum mechanics. Likewise, it is often effective in microscopical objects like atoms and molecules to treat quantum dynamics as a perturbation of classical motion.

Chapter 2
Dynamical Systems

Ocean snails (Left: Oliva porphyria, Right: Conus marmoreus). Each time, a photograph on the left, and a simulation by a dynamical system on the right.[1]

Dynamics can be viewed under different aspects, and with a variety of additional structures; accordingly, there are different definitions of dynamical systems. Whereas we will mainly consider *topological* and *differentiable* dynamical systems in this chapter, we will start with some larger generality.

It is only in Chapter 6 that *Hamiltonian* dynamical systems, which dominate in classical mechanics, will move center-stage. These are solutions of a special kind of differential equations of first order. They have a time-invariant measure, which, in the case of a vector space as the phase space, is the Lebesgue measure.

Chapter 9 is dedicated to such *measure-preserving* (not necessarily Hamiltonian) dynamical systems.

---

[1]The photos are from Chapter 10 (written by D. FOWLER and P. PRUSINKIEWICZ [FP]) in H. MEINHARDT: The Algorithmic Beauty of Sea Shells. 4. ed., c Springer 2009. Photos: courtesy of D. FOWLER and P. PRUSINKIEWICZ

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_2

The textbooks by KATOK and HASSELBLATT [KH] and by ROBINSON [Ro] consider general (not necessarily Hamiltonian) dynamical systems, as well as iterated non-invertible mappings (which are also often viewed as dynamical systems). ♢

2.1 Iterated Mappings, Dynamical Systems

"Nous devons donc envisager l'état présent de l'univers, comme l'effet de son état antérieur, et comme la cause de celui qui va suivre. Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée, et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome: rien ne serait incertain pour elle, et l'avenir comme le passé, serait présent à ses yeux."
PIERRE- SIMON LAPLACE *(1814), [Lap], page 2.*[2]

Given a mapping $f : M \to M$ of a set M into itself, we can *iterate* it by defining the *iterated map*

$$f^{(0)} := \mathrm{Id}_M \quad , \quad f^{(t)} := f \circ f^{(t-1)} \qquad (t \in \mathbb{N}).$$

For each $m \in M$, we obtain a sequence $a : \mathbb{N}_0 \to M$, $a_t = f^{(t)}(m)$. We then call M the phase space, t the time parameter, and m the initial point. We have not assumed f to be injective. So one can see from the graph of f that different initial points may have the same future.

2.2 Example For $f : \mathbb{N}_0 \to \mathbb{N}_0, m \mapsto \lfloor m/2 \rfloor$, the mapping yields the graph depicted in Figure 2.1.1 (left) by arrows from m to $f(m)$. ♢

Simple iterated mappings can already lead to very difficult questions.

2.3 Example (**Collatz Conjecture**) On the phase space $M := \mathbb{N}$, the mapping f is defined by $f(m) := m/2$ when m is even, and $f(m) := 3m + 1$ when m is odd. For instance, the sequence with initial point 7 begins as 7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1, 4, … and becomes cyclic once it reaches 1, as seen in Figure 2.1.1 (right). The conjecture by Lothar Collatz that *every* sequence contains 1 has been neither proved nor disproved since 1937. ♢

[2] Translation: "We therefore must consider the present state of the universe as the consequence of its earlier and the cause of its future states. An intelligence that would know, for some given moment, all forces that animate nature, as well as the respective positions of the elements of which it consists, and which would be comprehensive enough to subject these quantities to analysis, would in one and the same formula comprise the movements of the largest celestial bodies and the lightest of atoms; nothing would be uncertain to that intelligence, and future and past would be visible to it."

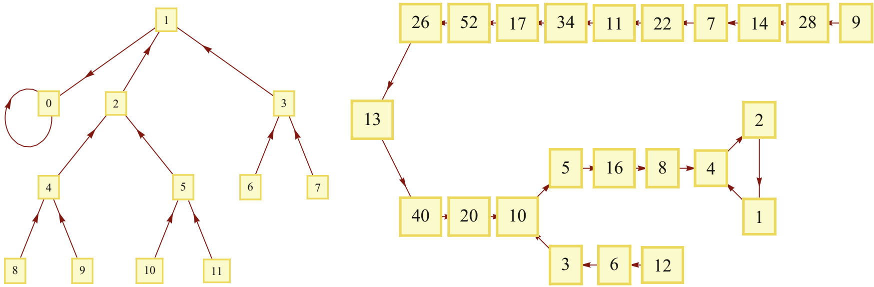
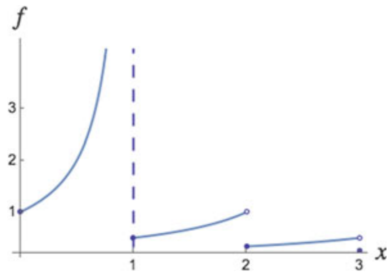


Figure 2.1.1 Iterated maps. Left: Example 2.2. Right: Collatz graph from Example 2.3

2.4 Example (**Calkin-Wilf Sequence**) For the decomposition $x = \lfloor x \rfloor + \{x\}$ of $x \in \mathbb{R}$, with $\lfloor x \rfloor \in \mathbb{Z}$ and $\{x\} \in [0, 1)$, we iterate the mapping

$$f : \mathbb{R}^+ \to \mathbb{R}^+ \quad , \quad x \mapsto \frac{1}{\lfloor x \rfloor + 1 - \{x\}},$$



beginning with 1. The first terms in the sequence are therefore $1, \frac{1}{2}, 2, \frac{1}{3}, \frac{3}{2}, \frac{2}{3}, 3, \frac{1}{4}, \frac{4}{3}, \frac{3}{5}, \frac{5}{2}, \frac{2}{5}, \frac{5}{3}, \frac{3}{4}, \ldots$ This sequence enumerates the positive rational numbers. A proof can be found in Calkin and Wilf [CW]. ◊

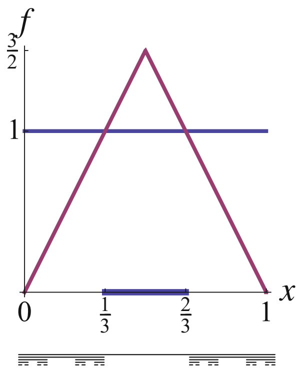
2.5 Exercise (**Cantor Set**)
The interval $I := [0, 1]$ contains the *hole* $L := (1/3, 2/3)$. We iterate the mapping

$$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \tfrac{3}{2}\big(1 - 2\,|x - 1/2|\big)$$

until we reach the hole. So we consider, for the start value $x_0 \in I$, the sequence $(x_n)_{n \in \mathbb{N}}$, $x_{n+1} := f(x_n)$. Show that the set

$$C := \{x_0 \in I \setminus L \mid \forall n \in \mathbb{N} : x_n \in I \setminus L\}$$

which consists of initial values that will never fall into the hole, is the Cantor-1/3 set. ◊



2.6 Remark (**Invertibility**) The fundamental equations of nature do not prescribe a direction of time. This is why non-bijective mappings defining dynamics play a role for events in physics only as coarse models. In classical mechanics, it is dynamical systems in the sense of the following definition, rather than general iterated mappings, that are in the center of interest.

Nevertheless we will occasionally return to non-injective iterated mappings (as the *logistic family* of Example 2.26), because these examples can feature complicated dynamics already in dimension one and are therefore good for the purpose of illustration. ♢

2.7 Definition *For the groups* $G := (\mathbb{Z}, +)$ *or* $G := (\mathbb{R}, +)$ *respectively, a family of mappings* $\Phi_t : M \to M \quad (t \in G)$ *on a set* M *is called a* **dynamical system***, if the following conditions are satisfied:*

$$\Phi_0 = \mathrm{Id}_M \quad \textit{and} \quad \Phi_{t_2} \circ \Phi_{t_1} = \Phi_{t_1+t_2} \quad (t_1, t_2 \in G). \tag{2.1.1}$$

Then M *is called* **phase space***. For* $G = \mathbb{Z}$*, the dynamical system is called* **discrete***, for* $G = \mathbb{R}$*, it is called* **continuous** *or a* **flow on** M*.*

2.8 Example (**Discrete Dynamical System**)
$(\Phi_t)_{t\in\mathbb{Z}}$ is a dynamical system if and only if $f := \Phi_1 : M \to M$ is bijective, with $\Phi_t = f^{(t)}$ and $\Phi_{-t} = (f^{-1})^{(t)}$ for $t \in \mathbb{N}_0$. ♢

2.9 Remarks

1. The phase space point $\Phi(t, m)$ describes the state of the dynamical system (e.g., the positions and velocities of the celestial bodies under consideration) at time t, if the state at time 0 was m.
2. The requirement $\Phi_0(m) = m$ is obviously justified, and it ensures that the mappings Φ_t are bijective.
   The requirement $\Phi_{t_2} \circ \Phi_{t_1} = \Phi_{t_1+t_2}$ of invariance under time translations will be satisfied for $G = \mathbb{R}$ when Φ is the solution of an autonomous system of differential equations (see Definition 3.13). For instance, in order to calculate the position and velocity of the earth in 10 months, we can first determine its state in 7 months, and then take that point in time as a new origin of the time axis and calculate the state 3 months later.
   However, for solutions Φ of differential equations with an explicit time dependence, this requirement will not be met. Let us imagine, for instance, that eight months from now, a fast comet will deflect the earth and let $\Phi(t, m)$ be the state of the earth at time t. Then *this* mapping Φ will not satisfy the above condition. Only after including the dynamics of the comet by enlarging the phase space to include its position and velocity do we again get an autonomous system of differential equations, and a dynamical system as its solution. ♢

2.10 Definition *For a dynamical system* $(\Phi_t)_{t\in G}$ *on the phase space* M *and a point* $m \in M$*, we call*

- *The mapping* $G \to M$, $t \mapsto \Phi_t(m)$ *the* **orbit curve through** m*.*
- *Its image* $\mathcal{O}(m) := \{\Phi_t(m) \mid t \in G\}$ *the* **orbit** *or* **trajectory through** m*.*
- m *is called* **equilibrium** *or* ***fixed point*** *if* $\mathcal{O}(m) = \{m\}$*.*
- $m \in M$ *is called* **periodic** *(or* ***closed****) with* **period** $T \in G$ *if* $T > 0$ *and* $\Phi_T(m) = m$*.*
  Here, T *is called* **minimal period** *if* $\Phi_t(m) \neq m$ *for all* $t \in (0, T)$*.*

Analogous definitions apply to non-invertible mappings, if one replaces the group $\mathbb{Z}$ with the set $\mathbb{N}_0$.

2.11 Example (**Matrix Powers as Dynamical Systems**) For an invertible symmetric matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ and $M := \mathbb{R}^n$, the mapping

$$\Phi : \mathbb{Z} \times M \to M \quad , \quad \Phi_t(x) \ := \ A^t x \quad \text{with } A^t \ = \ \sum\nolimits_{\lambda \in \mathrm{spec}(A)} \lambda^t P_\lambda ,$$

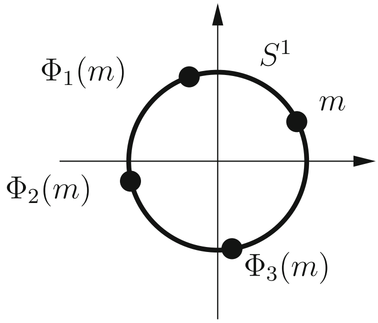
where $A = \sum_{\lambda \in \mathrm{spec}(A)} \lambda P_\lambda$ is the spectral representation of A, is a dynamical system. If $1 \notin \mathrm{spec}(A)$, i.e., is not an eigenvalue, then $x = 0$ is the only equilibrium. ◊

2.12 Exercises (**Period**)

1. On the phase space $S^1 = \{c \in \mathbb{C} \mid |c| = 1\}$, we define, with a parameter $\alpha \in \mathbb{R}$, the rotation by multiples of $2\pi\alpha$ by

$$\Phi_t : S^1 \to S^1 \ , \ \Phi_t(m) := \exp(2\pi i t \alpha)\, m \qquad (t \in \mathbb{Z}), \tag{2.1.2}$$

see the figure. Show:



   (a) The mappings (2.1.2) are a dynamical system.
   (b) For rational parameters $\alpha \in \mathbb{Q}$, i.e., $\alpha = \frac{q}{p}, q \in \mathbb{Z}, p \in \mathbb{N}$, and q, p relatively prime, every point in the phase space is periodic, with minimal period p.
   (c) For irrational $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, no point in the phase space is periodic.
2. The mapping $\mathbb{C} \to \mathbb{C}, z \mapsto z^m$ with $m \in \mathbb{N} \setminus \{1\}$ induces, by restriction to $|z| = 1$, a non-invertible mapping $f_m : S^1 \to S^1$ of the circle onto itself.
   Calculate, for $n \in \mathbb{N}$, the number $P_n(f_m)$ of periodic points of f_m with period n (where n doesn't have to be the minimal period), and show that the set of periodic points of f_m is dense in S^1. ◊

2.13 Theorem (**Orbits of Dynamical Systems**)

1. *The relation $m_1 \sim m_2$ if $m_2 \in \mathcal{O}(m_1)$, in words, 'if m_2 belongs to the orbit through m_1', is an equivalence relation on M.*
2. *If m is periodic with period T, then the same is true for all points of the orbit $\mathcal{O}(m)$; we then call this a* **periodic orbit**.
3. *If the orbit $\mathcal{O}$ has a minimal period $T > 0$, then this minimal period is unique, and the periods of $\mathcal{O}$ are the set $T\mathbb{N} = \{Tn \mid n \in \mathbb{N}\}$.*

Proof:

1. - Since $\Phi_0 = \mathrm{Id}_M$, the relation $\sim$ is *reflexive* (i.e., $m \sim m$ for all $m \in M$).
   - If $m_2 = \Phi_t(m_1)$, then $\Phi_{-t}(m_2) = \Phi_{-t} \circ \Phi_t(m_1) = \Phi_{t-t}(m_1) = m_1$; therefore the relation is *symmetric*.
   - With $m_2 = \Phi_{t_1}(m_1)$ and $m_3 = \Phi_{t_2}(m_2)$, we conclude $m_3 = \Phi_{t_1+t_2}(m_1)$; so $\sim$ is *transitive*.
2. If $m' = \Phi_t(m)$, then $\Phi_T(m') = \Phi_{T+t}(m) = \Phi_t \circ \Phi_T(m) = \Phi_t(m) = m'$; so T is also a period of m'.
3. If $S > 0$ is also a minimal period of $\mathcal{O}$, then $T \notin (0, S)$, hence $T \geq S$; conversely it is also true that $S \geq T$.

   The latter is true for all periods S. If it were true that $S \notin T\mathbb{N}$, one would have a unique representation $S = nT + r$ with $n \in \mathbb{N}$ and $r \in (0, T)$. Then r would also be a period of $m \in \mathcal{O}$:

$$\Phi_r(m) = \Phi_{S-nT}(m) = \Phi_S \circ \Phi_{-nT}(m) = \Phi_S(m) = m.$$

   This would be in contradiction to the minimality of T. □

For the group $G = \mathbb{Z}$ of times, the equilibria have a minimal period of 1, whereas for $G = \mathbb{R}$, they do not have a minimal period.

2.14 Definition *A subset $N \subseteq M$ of the phase space M is called*

- **forward invariant**, *if $\Phi_t(N) \subseteq N$ for all $t \in G$ with $t \geq 0$.*
- **invariant**, *if $\Phi_t(N) \subseteq N$ for all $t \in G$.*

Invariant sets have the property that $\Phi_t(N) = N$ for all $t \in G$, since $\Phi_{-t}(N) \subseteq N$ implies $N = \Phi_t\big(\Phi_{-t}(N)\big) \subseteq \Phi_t(N)$.

We call a nonempty invariant subset N of M *minimal* (in the set theoretic sense), if it does not contain a strict subset that is invariant. The minimal invariant subsets are precisely the orbits. So we have understood a dynamical system $(\Phi_t)_{t\in G}$ once we know its orbits and the restriction of $\Phi_t : M \to M$ to these orbits.

2.15 Exercise (**Minimal Period of a Dynamical System**)
We call $T > 0$ **period of the dynamical system**, if $\Phi_T = \mathrm{Id}_M$.

Show that a discrete dynamical system $\Phi : \mathbb{Z} \times M \to M$ on a finite set $M \neq \emptyset$ has the least common multiple of the minimal periods of its orbits as its minimal period. ◊

2.2 Continuous Dynamical Systems

The number of questions about a dynamical system increases enormously once a topology is available (see Appendix A.1), so that we can, for instance, talk about limits.

2.16 Definition *A dynamical system* $(\Phi_t)_{t\in G}$ *on the phase space* M *is called a* **continuous** *or* **topological** *dynamical system, if* M *is a topological Hausdorff space and*

$$\Phi : G \times M \to M \quad , \quad \Phi(t, m) := \Phi_t(m)$$

is continuous.[3]

2.17 Remarks (**Topological Dynamical Systems**)

1. As $\mathbb{Z}$ is a discrete topological space, i.e., every subset is open, the continuity of Φ for $G = \mathbb{Z}$ is equivalent to the continuity of Φ_t for each $t \in \mathbb{Z}$. This in turn is equivalent to $\Phi_1 : M \to M$ being a homeomorphism (see Definition A.17). Conversely, every homeomorphism $f : M \to M$ generates a continuous dynamical system by iteration.
2. In almost all applications, the phase space of a dynamical system is a Hausdorff space in a natural way. However, one needs to decide on a case-by-case basis whether the dynamics is continuous. For instance, in the case of billiard,[4] or if point masses collide, this is not always the case.
3. If one omits topology, and hence the requirement of continuity of Φ with it, then one loses the control that for finite time t, the state $\Phi(t, m')$ stays close to the state $\Phi(t, m)$ if the initial points m and m' are close to each other. This is however a property of practical interest, since in an experiment, m can only be measured with finite precision.
4. In all the introductory examples, M is the space of positions q and velocities v of the massive bodies under consideration. For instance, in the case of earth in 3-space, M is the topological space $\mathbb{R}^3 \times \mathbb{R}^3$, and $m = (q, v)$. The value $\Phi(t, m)$ gives the state (here: position and velocity) of the body after time t, given that its state at time 0 was m.
   However, we also want to consider, for example, the motion of a bead on a circular wire. Its position there is given as a point on the circle $S^1 := \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$, and the phase space is the manifold $M := S^1 \times \mathbb{R}$ (where $\mathbb{R}$ is for the velocity of the circular motion). The notion of a manifold is introduced in Appendix A.2.
5. There also occur phase spaces that are not manifolds.
   For instance, the *Schrödinger equation* from quantum mechanics,

$$\frac{\mathrm{d}\Phi_t}{\mathrm{d}t} = -i\, H\, \Phi_t \quad , \quad \Phi_0 = \mathbb{1}_{\mathcal{H}},$$

   with a self-adjoint operator $H = H^* : \mathcal{H} \to \mathcal{H}$ defines a unitary time evolution $\Phi_t = \exp(-iHt)$ with $t \in \mathbb{R}$ on the $\mathbb{C}$-Hilbert space $(\mathcal{H}, \langle\cdot, \cdot\rangle)$. This latter is a topological space (being a normed vector space), and the mapping

---

[3] Here, $(\mathbb{R}, +)$ or $(\mathbb{Z}, +)$ respectively are considered as a topological group (Definition E.16), and $G \times M$ carries the product topology (Appendix A.1).

[4] A mathematical variation of the game of pool.

$$\Phi : \mathbb{R} \times \mathcal{H} \to \mathcal{H} \quad , \quad \Phi(t, \psi) = \Phi_t(\psi)$$

is continuous.[5]

6. Part of our task will be to determine Φ: for the group $G = \mathbb{R}$ by solving ordinary differential equations; and for $G = \mathbb{Z}$ by iterating a mapping. In this process, the mappings Φ will turn out not only to be continuous, but also arbitrarily often differentiable ('smooth'), provided the same is true for the differential equation or, respectively, the mapping being iterated.
7. If in Definition 2.16, one admits arbitrary topological groups[6] $(G, \circ)$, requiring more generally that $\Phi_{t_2} \circ \Phi_{t_1} = \Phi_{t_2 \circ t_1}$, one obtains the notion of a (topological) *group action*, more precisely *action, or operation, from the left*, in contradistinction to an *action from the right*, which requires $\Phi_{t_2} \circ \Phi_{t_1} = \Phi_{t_1 \circ t_2}$. Such group actions can occur for instance as symmetries of a dynamical system. The special case of dynamical systems is simpler inasmuch as the groups $\mathbb{R}$ and $\mathbb{Z}$ are abelian. However, they are not compact, and this makes the analysis more difficult. ◊

2.18 Examples (**Topological Dynamical Systems**)

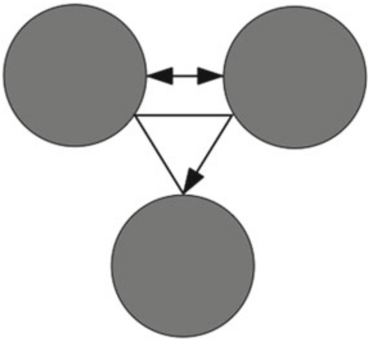
1. One simple example is the *free movement* of a celestial body in vacuum. The phase space is again $M := \mathbb{R}^3_q \times \mathbb{R}^3_v$. As, by hypothesis, no forces act on the body, its acceleration is $\frac{\mathrm{d}^2}{\mathrm{d}t^2} q = 0$. So the velocity, i.e., the derivative $v = \frac{\mathrm{d}}{\mathrm{d}t} q$ of the position with respect to time, is constant in time: $\frac{\mathrm{d}}{\mathrm{d}t} v = 0$, and we have the solution
$$\Phi(t, m) = (q + v\,t,\ v) \qquad \big(t \in \mathbb{R},\ m = (q, v) \in M\big)$$
continuous in the initial data m and time t.
2. In contrast, the example of the two body problem of celestial mechanics, given in the introduction, is strictly speaking not even a dynamical system in the sense of Definition 2.7, because both bodies can collide in finite time. However, if the phase space $(\mathbb{R}^3_q \setminus \{0\}) \times \mathbb{R}^3_v$ is restricted by the (flow-invariant) condition $q \times v \neq 0$, i.e., non-vanishing angular momentum, then one obtains a continuous dynamical system.
3. For a finite set $\mathcal{A} \neq \emptyset$, a discrete dynamics can be defined on the sequence space
$$M := \mathcal{A}^{\mathbb{Z}} = \{a : \mathbb{Z} \to \mathcal{A}\} = \{(a_k)_{k \in \mathbb{Z}} \mid a_k \in \mathcal{A}\}$$
by setting
$$\Phi_t : M \to M \quad , \quad \big(\Phi_t(a)\big)_k := a_{k+t} \qquad (t \in \mathbb{Z}). \tag{2.2.1}$$
If the set $\mathcal{A}$, called *alphabet*, is equipped with the discrete topology, and the *sequence space* M (also known as *shift space*) with the product topology, then M is a topological Hausdorff space, and the *shift* Φ is a continuous dynamical system. By Tychonoff's theorem (Theorem A.19), the phase space M is compact.

[5] But it is only when $\dim(\mathcal{H}) < \infty$ that the phase space will be a (finite-dimensional) manifold according to the definition A.25.

[6] See Appendix E.1 and E.2.

Even though the phase space of this dynamical system is not a manifold (assuming $|\mathcal{A}| > 1$), it can nevertheless be used for the analysis of physically relevant chaotic dynamical systems. An easy example is the one of a particle, reflected elastically by three discs, see figure. The orbits that stay bounded in the past and the future have infinitely many such collisions. As they cannot collide twice with the same disc in succession, their collision sequence is a point in

$$\Sigma := \left\{a \in \mathcal{A}^{\mathbb{Z}} \mid \forall k \in \mathbb{Z} : a_{k+1} \neq a_k\right\},$$



for the alphabet $\mathcal{A} := \{1, 2, 3\}$ enumerating the discs. It turns out that there is a bijection between Σ and the phase space points on the discs belonging to these bounded orbits. There are two periodic orbits shown in the figure, with periodic symbol sequences $\overline{12}$ and $\overline{123}$, respectively.
See [KS, Section 3] for more details. ◊

2.19 Exercise (Shift)
As in Example 2.18.3, let $\mathcal{A}$ be an alphabet and $M := \mathcal{A}^{\mathbb{Z}}$.

(a) Check that $d : M \times M \to [0, \infty)$, given by

$$d(x, y) := \sum_{j \in \mathbb{Z}} 2^{-|j|}\, d_{\mathcal{A}}(x_j, y_j) \qquad \left(x = (x_j)_{j\in\mathbb{Z}},\ y = (y_j)_{j\in\mathbb{Z}} \in M\right),$$

with $d_{\mathcal{A}}(a, b) := 0$ if $a = b$ and $d_{\mathcal{A}}(a, b) := 1$ otherwise, defines a metric on M. Show that the shift Φ is continuous.

(b) How many periodic points $m \in M$, and how many periodic orbits with period $n \in \{2, 3, 4\}$, does Φ from (2.2.1) have, when the alphabet is $\mathcal{A} = \{0, 1\}$? Specify all points with minimal period n for $n \in \{2, 3, 4\}$.

(c) Show that there exists an $x \in M$ whose orbit is dense in M, i.e.,

$$\overline{\{\Phi_t(x) \mid t \in \mathbb{Z}\}} = M.$$

Remark: This means that the continuous dynamical system $\Phi : \mathbb{Z} \times M \to M$ is *topologically transitive*, i.e., for open non-empty sets $A, B \subseteq M$, there exists a $t \in \mathbb{Z}$ such that $\Phi_t(A) \cap B \neq \emptyset$. ◊

The long-term behavior of a point in phase space M is described (at least in case M is compact) by its limit sets:

2.20 Definition
For a continuous dynamical system $\Phi : G \times M \to M$ and $x \in M$, we call

$$\alpha(x) := \left\{y \in M \;\middle|\; \exists (t_n)_{n\in\mathbb{N}} \text{ with } \lim_{n\to\infty} t_n = -\infty \text{ and } \lim_{n\to\infty} \Phi(t_n, x) = y\right\},$$

$$\omega(x) := \left\{ y \in M \;\middle|\; \exists (t_n)_{n\in\mathbb{N}} \text{ with } \lim_{n\to\infty} t_n = +\infty \text{ and } \lim_{n\to\infty} \Phi(t_n, x) = y \right\}$$

the α-**limit set of** x *and the* ω-**limit set of** x *respectively.*

On the one hand, these sets are invariants of the orbit $\mathcal{O}(x)$, i.e., $\alpha(y) = \alpha(x)$ and $\omega(y) = \omega(x)$ if $y \in \mathcal{O}(x)$; on the other hand, these sets are invariant themselves.

Now one isn't merely interested in individual orbits, but also in the behavior of neighboring orbits. For instance, it is comforting that even in the case of the small change in the velocity of the earth, as caused by a meteorite impact, the new orbit will stay near the old one for all time. We first study the stability of fixed points, later we will study the stability of periodic orbits.

2.21 Definition (**Stability**) *Let* $m_0 \in M$ *be a fixed point of the continuous dynamical system* $\Phi : G \times M \to M$.

1. m_0 *is called* **Lyapunov-stable** *if for every neighborhood* $U \subseteq M$ *of* m_0*, there exists a (smaller) neighborhood* V *of* m_0*, such that for all* $t \geq 0$,
$$\Phi_t(V) \equiv \{\Phi_t(m) \mid m \in V\} \subseteq U.$$
2. *Otherwise* m_0 *is called* **unstable**.
3. m_0 *is called* **asymptotically stable** *if* m_0 *is Lyapunov-stable and there exists a forward-invariant neighborhood* $V \subseteq M$ *of* m_0*, with*
$$\lim_{t\to\infty} \Phi_t(m) = m_0 \qquad (m \in V).$$

2.22 Exercise (**Stability**)
On the phase space $M := \mathbb{C}$, consider, for a parameter $\lambda \in \mathbb{C} \setminus \{0\}$, the mappings $\Phi_t : M \to M$, $\Phi_t(m) := \lambda^t m$, with $t \in \mathbb{Z}$. Show:

(a) These mappings form a continuous dynamical system, and $0 \in M$ is a fixed point.
(b) This fixed point is Lyapunov-stable if and only if $|\lambda| \leq 1$.
(c) The fixed point is asymptotically stable if and only if $|\lambda| < 1$. $\Diamond$

2.23 Definition

- *A compact invariant subset* $A \subseteq M$ *is called* **attractor** *of the continuous dynamical system* $\Phi : G \times M \to M$ *if there exists an open neighborhood* $U_0 \subseteq M$ *of* A *such that*

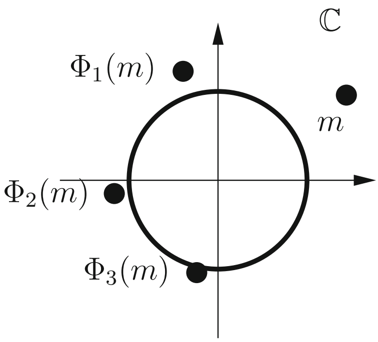
(a) U_0 *is forward invariant;*
(b) For every open neighborhood V *of* A *with* $A \subseteq V \subseteq U_0$*, there exists a* $\tau > 0$ *such that* $\Phi_t(U_0) \subseteq V$ *for all* $t \geq \tau$.

- *The* **basin** *of an attractor* A *is the union of all open neighborhoods* U_0 *of* A *that satisfy items (a) and (b).*

As a consequence, that basin B itself is an open neighborhood of A satisfying property (a). However, as shown by the next example, property (b) is not in general satisfied by B.

2.24 Example (**Attractor**) On the phase space $\mathbb{C}$, a continuous dynamical system $\Phi : \mathbb{Z} \times \mathbb{C} \to \mathbb{C}$ with a parameter $\lambda \in \mathbb{R}$ is given by iteration of the homeomorphism

$$\Phi_1(m) := \begin{cases} \frac{e^{i\lambda} m}{\sqrt{|m|}} & , m \neq 0 \\ 0 & , m = 0, \end{cases}$$



see the figure. Here, $A := S^1 \subset \mathbb{C}$ is an attractor, and its basin is $\mathbb{C} \setminus \{0\}$. This is because $\Phi_t : \mathbb{C} \to \mathbb{C}$ maps circles of radius $r > 0$ onto circles of radius $r^{(2^{-t})}$. So the circle S^1 (which is compact) is invariant, and the images of the open annuli

$$U(r_1, r_2) \;:=\; \{c \in \mathbb{C} \mid |c| \in (r_1, r_2)\} \;\subset\; \mathbb{C} \setminus \{0\}$$

converge, for $0 < r_1 < 1 < r_2$, to A in the following sense:

For every $c > 1$, there exists $\tau \in \mathbb{N}$ with $\Phi_t\big(U(r_1, r_2)\big) \subseteq U(1/c, c)$ if $t \geq \tau$. Due to the compactness of A, every open neighborhood V of A contains some $U(1/c, c)$. The basin of A cannot contain the fixed point $0 \in \mathbb{C}$, but does contain all the sets $U(r_1, r_2)$; it is therefore equal to $\mathbb{C} \setminus \{0\}$. ◊

2.25 Exercise (**Attractor**)
Let $\Phi : G \times M \to M$ be a continuous dynamical system.

(a) Is the union of two attractors again an attractor?
(b) Show that for an attractor $A \subseteq M$ and a corresponding set U_0 (from Definition 2.23 of an attractor), one has $A = \bigcap_{t \geq 0} \Phi_t(U_0)$. ◊

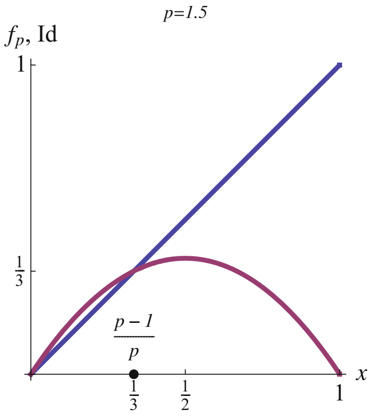
2.26 Example (**Logistic Family**)
For a parameter $p \in [0, 4]$, we consider the (non-invertible) *logistic mapping* on phase space $M := [0, 1]$

$$f_p : M \to M \quad , \quad f_p(x) := p\,x\,(1 - x). \tag{2.2.2}$$

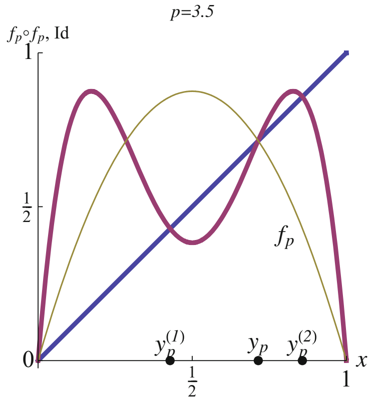
- The point $0 \in M$ is mapped by f_p into itself, i.e., is a fixed point, for every value of the parameter p.
- If $p \leq 1$, then $f_p(x) < x$ for every $x > 0$, and so it follows for all $m \in M$ that

$$\lim_{t \to \infty} f_p^{(t)}(m) = 0.$$

- Moreover, for the parameter $p \in (1, 4]$, there exists a second fixed point of f_p in M, namely $y_p := \frac{p-1}{p}$, see the figure on the right.



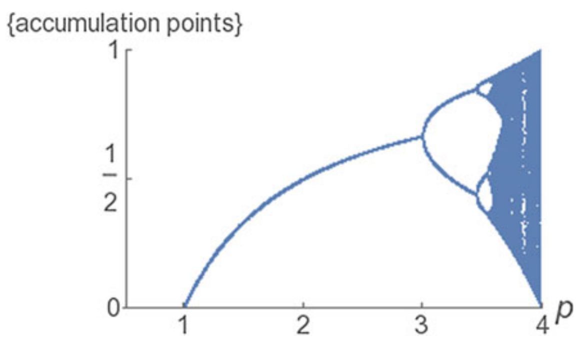
- If $p \in (1, 3]$, all sequences with initial point $m \in (0, 1)$ satisfy $\lim_{t\to\infty} f_p^{(t)}(m) = y_p$. For $p \in (1, 2]$, this can be seen as follows: we only need to consider initial values $x \in (0, 1/2]$ since f_p maps the right half interval $[1/2, 1)$ into the left half $(0, 1/2]$.
  For $x \in (0, y_p)$, one has $f_p(x) \in (x, y_p)$; for $x \in (y_p, 1/2]$, one has $f_p(x) \in (y_p, x)$.
  The general case is to be studied in Exercise 2.27.
- For parameter values $p \in (3, 4]$, the iterated map $f_p \circ f_p$ has four fixed points. In the figure on the right, they can be found as intersections of the diagonal and the graph of $f_p \circ f_p$. Two of these fixed points are the fixed points of f_p that we discussed already. The other two, let's call them $y_p^{(1)}$ and $y_p^{(2)}$, are mapped into each other under the logistic mapping, i.e., $f_p\left(y_p^{(1)}\right) = y_p^{(2)}$ and $f_p\left(y_p^{(2)}\right) = y_p^{(1)}$.



Due to the Bolzano-Weierstrass theorem, we know that the sequences $t \mapsto f_p^{(t)}(m)$ will always have an accumulation point.

We are interested in the set of accumulation points, in dependence on the parameter p and the initial value m.

As $f_p(0) = 0$, there is only the accumulation point 0 in case $m = 0$. For typical initial points m, however, a complicated structure of accumulation points arises in dependence on the parameter p, as seen in the figure on the right (with initial value $m = 0.01$). Let us note as an aside that the iterated logistic map is used in physics as a simple model for the transition from laminar to turbulent flows in fluids, where large values of p are associated with the turbulent regime, see FEIGENBAUM [Fei]. ◊



2.27 Exercises (**Logistic Family**)

1. We consider the n^{th} iterate $f_4^{(n)}$ of the logistic mapping (2.2.2).
   Show that $f_4^{(n)}$ has exactly 2^n fixed points in $[0, 1]$, by analyzing the intervals of monotonicity of $f_4^{(n)}$.
2. Show that for parameter values $p \in (1, 3)$, the logistic mapping (2.2.2) has the fixed point $y_p = (p-1)/p$, and that $\lim_{n\to\infty} f_p^{(n)}(x) = y_p$ for all $x \in (0, 1)$.
   Hint: What are the values of $f_p'(y_p)$ in the interval $1 \leq p \leq 3$? ◊

We now want to compare continuous dynamical systems with each other.

2.28 Definition *For two continuous dynamical systems* $\Phi^{(i)} : G \times M^{(i)} \to M^{(i)}$ *with* $i = 1, 2$,

- *we call* $\Phi^{(2)}$ *a (topological)* **factor of** $\Phi^{(1)}$, *and* $\Phi^{(2)}$ **semiconjugate** *to* $\Phi^{(1)}$, *if there exists a continuous surjection* $h : M^{(1)} \to M^{(2)}$ *with* $\Phi_t^{(2)} \circ h = h \circ \Phi_t^{(1)}$ *for all* $t \in G$, *i.e., if the following diagram commutes:*

$$\begin{array}{ccc} M^{(1)} & \xrightarrow{\Phi_t^{(1)}} & M^{(1)} \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h} \\ M^{(2)} & \xrightarrow{\Phi_t^{(2)}} & M^{(2)} \end{array} \tag{2.2.3}$$

- *we call* $\Phi^{(2)}$ **conjugate to** $\Phi^{(1)}$, *if* $\Phi^{(2)}$ *is a factor of* $\Phi^{(1)}$, *such that the diagram (2.2.3) is even valid with a homeomorphism* $h : M^{(1)} \to M^{(2)}$. *In this case, h is called a* **conjugacy**.

2.29 Remarks (**Conjugacy**)

1. As inverses and compositions of homeomorphisms are again homeomorphisms, the definition of conjugacy is independent of the numbering of the dynamical systems, and we obtain a categorization into classes of mutually conjugate continuous dynamical systems.
2. If two continuous dynamical systems are conjugate at all, there are usually many conjugacies. For if h from (2.2.3) is a conjugacy, then, for example, all $h_s := h \circ \Phi_s^{(1)}$ with $s \in G$ are conjugacies, too; they are different from h if $\Phi_s^{(1)} \neq \mathrm{Id}_{M^{(1)}}$.
3. In the proof of Theorem 2.31 on circle rotations, a semiconjugacy will be used as a proof technique. ◊

As the notions defined in this chapter are purely of a topological nature, they carry over to conjugate systems. In particular, one obtains:

2.30 Exercise (**Conjugacy**) Let $h : M^{(1)} \to M^{(2)}$ be a conjugacy of the continuous dynamical systems $\Phi^{(i)} : G \times M^{(i)} \to M^{(i)}$. Prove:

(a) $x_1 \in M^{(1)}$ is an equilibrium of $\Phi^{(1)}$ if and only if $x_2 := h(x_1) \in M^{(2)}$ is an equilibrium of $\Phi^{(2)}$. Conjugate equilibria do not differ with regard to Lyapunov-stability or asymptotic stability.

(b) The $\Phi^{(1)}$-orbit $\mathcal{O}(x_1)$ through $x_1 \in M^{(1)}$ is periodic if and only if the $\Phi^{(2)}$-orbit $\mathcal{O}(x_2)$ through $x_2 := h(x_1) \in M^{(2)}$ is periodic. In that case, the periods are equal.
(c) The image of the ω-limit set $\omega(x_1)$ of $x_1 \in M^{(1)}$ equals

$$h\big(\omega(x_1)\big) = \omega\big(h(x_1)\big). \qquad \diamondsuit$$

In order to show that two continuous dynamical systems are *not* conjugate, it suffices to use an invariant under conjugacy. For example, according to Exercise 2.30(b), the systems are not conjugate if a periodic orbit of a certain period exists in the first system, but not in the second.

In the proof of the following theorem, a quantity called *rotation number* is almost such an invariant.

Namely, we study when rotations of a circle from Exercise 2.12.1 are conjugate. For these dynamical systems

$$\Phi^{(\gamma)} : \mathbb{Z} \times S^1 \to S^1 \quad , \quad \Phi^{(\gamma)}(t, m) = \exp(2\pi i \gamma t)\, m \qquad (\gamma \in \mathbb{R})$$

one even has: $\Phi^{(\alpha)} = \Phi^{(\beta)}$ if and only if $\alpha - \beta \in \mathbb{Z}$. So we may assume, without loss of generality, that $\gamma \in [0, 1)$.

2.31 Theorem (Circle Rotations) *Two such circle rotations $\Phi^{(\alpha)}$ and $\Phi^{(\beta)}$ are conjugate if and only if $\alpha = \beta$ or $\alpha = 1 - \beta$.*

Proof:

- For $\alpha = \beta$, we can take Id_{S^1} as a conjugacy, for $\alpha = 1 - \beta$, we can take the mapping $S^1 \to S^1, z \mapsto \bar{z}$.
- If, conversely, $\Phi_t^{(\beta)} = h \circ \Phi_t^{(\alpha)} \circ h^{-1}$ $(t \in \mathbb{Z})$ for a homeomorphism $h : S^1 \to S^1$, then we will *lift* these dynamical systems to the phase space $\mathbb{R}$. This process is explained in the following: The mapping

$$\pi : \mathbb{R} \to S^1 \quad , \quad x \mapsto \exp(2\pi i x)$$

is continuous and, intuitively speaking, it rolls up the real line onto the circle. π is a group homomorphism from $(\mathbb{R}, +)$ to $(S^1, \cdot)$, since $\pi(x + y) = \pi(x)\pi(y)$ by the functional equation of exp.
- We call a dynamical system

$$\tilde{\Phi}^{(\gamma)} : \mathbb{Z} \times \mathbb{R} \to \mathbb{R}$$

a π-*lift* of the circle rotation $\Phi^{(\gamma)} : \mathbb{Z} \times S^1 \to S^1$, if $\pi \circ \tilde{\Phi}_t^{(\gamma)} = \Phi_t^{(\gamma)} \circ \pi$ for all $t \in \mathbb{Z}$, i.e.,

$$\exp\left(2\pi i \tilde{\Phi}_1^{(\gamma)}(x)\right) = \exp\big(2\pi i (x + \gamma)\big) \qquad (x \in \mathbb{R}),$$

i.e., $\tilde{\Phi}_1^{(\gamma)}(x) = x + \gamma - n_\gamma$ for a certain $n_\gamma \in \mathbb{Z}$. (It is by continuity of $\tilde{\Phi}^{(\gamma)}$ that n_γ does not depend on x.)

- Similarly, we call a continuous mapping $\tilde{h} : \mathbb{R} \to \mathbb{R}$ a π-*lift* of the conjugacy $h : S^1 \to S^1$ if $\pi \circ \tilde{h} = h \circ \pi$, i.e., $\exp\big(2\pi i \tilde{h}(x)\big) = h\big(\exp(2\pi i x)\big)$.
  Then it must be true that $\tilde{h}(x+1) = \tilde{h}(x) + n$ for some $n \in \mathbb{Z}$. As $\tilde{h}$ is strictly monotonic, we have $n \neq 0$. On the other hand, since there is no $y \in (x, x+1)$ with $\tilde{h}(y) - \tilde{h}(x) \in \mathbb{Z}$, the only options for n are $n = -1$ and $n = 1$.
- The rotation number of the π-lift $\tilde{\Phi}^{(\gamma)}$ is defined as

$$R(\gamma) := \lim_{t\to\infty} \frac{\tilde{\Phi}^{(\gamma)}(t, x)}{t} \tag{2.2.4}$$

  and is indeed independent of the starting point $x \in \mathbb{R}$; namely we have $R(\gamma) = \gamma - n_\gamma$. It makes sense to choose $n_\gamma = 0$. Then, for dynamical systems $\Phi^{(\beta)}$ that are conjugate to $\Phi^{(\alpha)}$, the only possible rotation number is, dependent on whether $\tilde{h}$ is strictly increasing or decreasing,

$$R(\beta) = \lim_{t\to\infty} \frac{\tilde{h} \circ \Phi_t^{(\alpha)} \circ \tilde{h}^{-1}(x)}{t} \in \begin{cases} \alpha + \mathbb{Z}, & \tilde{h} \text{ increasing} \\ -\alpha + \mathbb{Z}, & \tilde{h} \text{ decreasing} \end{cases}.$$

  This shows that only $\beta = \alpha$ and $\beta = 1 - \alpha$ can be solutions. □

So most circle rotations are not topologically conjugate.

On the other hand, one can also define the rotation number $R(f)$ of the iterated mapping for other diffeomorphisms $f : S^1 \to S^1$ of the circle (see Def. 2.36) that are not rotations, in analogy to (2.2.4), and the following remarkable theorem (see Herman [Her]) applies:

2.32 Theorem (**Denjoy**) *If the rotation number $R(f)$ of a diffeomorphism $f \in C^2(S^1, S^1)$ is irrational, then the dynamical system defined by f is conjugate to the circle rotation $\Phi^{(R(f))}$.*

2.3 Differentiable Dynamical Systems

In order to use techniques from analysis for continuous dynamical systems, it is natural to assume that their phase space is a differentiable manifold. We give a systematic introduction to manifolds in Appendix A. Here, we will only consider the case of submanifolds of $\mathbb{R}^n$. This is not really a loss of generality (see Theorem A.4).

2.33 Definition
For an open subset $W \subseteq \mathbb{R}^n$ and $f \in C^1(W, \mathbb{R}^m)$, we call $y \in \mathbb{R}^m$ a **regular value** *of f if for all $q \in W$ with $f(q) = y$, the derivative $\mathrm{D}_q f : \mathbb{R}^n \to \mathbb{R}^m$ is surjective.*

This notion will first serve to define submanifolds of $\mathbb{R}^n$. It will be generalized in (A.45) to mappings between manifolds.

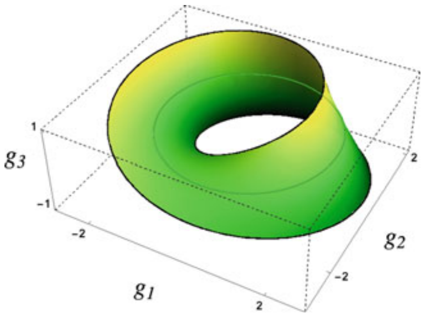
2.34 Definition
*For $p \in \{0, \ldots, n\}$, a subset $M \subseteq \mathbb{R}^n$ is called a p-***dimensional submanifold** *of $\mathbb{R}^n$ if there is, for each point $x \in M$, a neighborhood $V_x \subseteq \mathbb{R}^n$ such that one can write, with a suitable $f \in C^1\big(V_x, \mathbb{R}^{n-p}\big)$ having 0 as a regular value,*

$$M \cap V_x = f^{-1}(0).$$

In the simplest case, $M = f^{-1}(0)$; however, one also wants to admit as manifolds sets like the one in the following example:

2.35 Example (**Möbius Strip**)
For $U := \mathbb{R} \times (-1, 1)$ and $g \in C^\infty(U, \mathbb{R}^3)$ defined by

$$g(x, y) := \begin{pmatrix} \left(2 - y \sin \frac{x}{2}\right) \sin x \\ \left(2 - y \sin \frac{x}{2}\right) \cos x \\ y \cos \frac{x}{2} \end{pmatrix},$$



the set $M := g(U) \subset \mathbb{R}^3$ is called the *Möbius strip*.
The range of angles $x \in [0, 2\pi)$ would be sufficient to parametrize M: Whereas angles $x/2$ occur in the of g, one has $g(x + 2\pi, y) = g(x, -y)$. The set $g(\mathbb{R} \times \{0\})$ is a circle of radius 2. Since the surface M has only *one* side, it cannot be the level set $f^{-1}(0)$ of a regular value 0 for any function f; for otherwise $\nabla f(x) \neq 0$ would be perpendicular to the surface at each point $x \in M$ and thus distinguish one of two sides. ♢

2.36 Definition *Let $U \subseteq \mathbb{R}^n$ be open and $f \in C^1(U, \mathbb{R}^n)$.*

- *f is called a* **diffeomorphism onto the image** *$V := f(U) \subseteq \mathbb{R}^n$ if V is open, $f : U \to V$ is bijective, and $f^{-1} : V \to U$ is also continuously differentiable.*
- *f is called a* **local diffeomorphism** *if each point $x \in U$ has an open neighborhood $U_x \subseteq U$ such that the restriction $f\restriction_{U_x}$ is a diffeomorphism onto the image.*
- *For $r \in \mathbb{N}$ and open sets $U, V \subset \mathbb{R}^n$, a mapping $f \in C^r(U, V)$ is called a C^r-***diffeomorphism** *if f is a diffeomorphism onto the image V (and by consequence, the inverse mapping $f^{-1} \in C^r(V, U)$).*

One can view diffeomorphisms as changes of coordinates, and as one prefers to use coordinates that are adapted to the problem at hand, diffeomorphisms are a frequently used class of mappings.

2.37 Example (**Affine Maps**) An affine map $f : \mathbb{R}^n \to \mathbb{R}^n$ is of the form $f(x) = Ax + b$ where $A \in \mathrm{Mat}(n, \mathbb{R})$ and $b \in \mathbb{R}^n$. It is a diffeomorphism if and only if it is bijective, i.e., if and only if $A \in \mathrm{GL}(n, \mathbb{R})$. ♢

One can see from this example that the regularity of the Jacobi matrix $\mathrm{D}f$ influences the invertibility of the mapping f, for in this case one has $\mathrm{D}f = A$.

2.38 Theorem (**Local Diffeomorphisms**) *For an open set $U \subseteq \mathbb{R}^n$, a function $f \in C^1(U, \mathbb{R}^n)$ is a local diffeomorphism if and only if*

$$\mathrm{D}f(x) \in \mathrm{GL}(n, \mathbb{R}) \text{ for all } x \in U.$$

Proof:

- Let f be a local diffeomorphism, $x \in U$, and $g : V_x \to U_x$ the inverse function of the diffeomorphism $f\restriction_{U_x} : U_x \to V_x$. Then one has, by the chain rule,

$$\mathrm{D}g\big(f(x)\big)\,\mathrm{D}f(x) \;=\; \mathrm{D}(g \circ f)(x) \;=\; \mathrm{D}\,\mathrm{Id}_{U_x}(x) \;=\; \mathbb{1}, \text{ so } \mathrm{D}f(x) \in \mathrm{GL}(n, \mathbb{R}).$$

- Conversely, assume $\mathrm{D}f(x) \in \mathrm{GL}(n, \mathbb{R})$. To find the local inverse of f at $x \in U$, we apply the inverse function theorem. As we shall see, this in fact follows from an application of the implicit function theorem to

$$F : \mathbb{R}^n \times U \to \mathbb{R}^n \quad , \quad (y, z) \mapsto -y + f(z).$$

  By hypothesis, one has, with $X := \big(f(x), x\big)$, the relation

$$\mathrm{D}_2 F(X) = \mathrm{D}f(x) \in \mathrm{GL}(n, \mathbb{R}) \quad \text{, and} \quad F(X) = 0.$$

  Using the implicit function theorem gives the existence of open neighborhoods $V \subseteq \mathbb{R}^n$ of $f(x)$, $W \subseteq U$ of x and of a mapping $g \in C^1(V, W)$ such that $F(y, g(y)) = f(g(y)) - y = 0$ $(y \in V)$. We define $U_x := g(V) \subseteq W$. Both g and $f\restriction_{U_x}$ are injective, for otherwise $f \circ g = \mathrm{Id}_V$ would be impossible. But then we also have $g \circ f\restriction_{U_x} = \mathrm{Id}_{U_x}$, and by the chain rule, $\mathrm{D}g(y) \in \mathrm{GL}(n, \mathbb{R})$ for all $y \in V$. By the following theorem, this guarantees that U_x is an open neighborhood of x. □

2.39 Theorem *Let $U \subseteq \mathbb{R}^n$ be open and $f \in C^1(U, \mathbb{R}^n)$. If f is* **regular**, *i.e., if $\mathrm{D}f(x) \in \mathrm{GL}(n, \mathbb{R})$ for all $x \in U$, then $f(V)$ is open provided $V \subseteq U$ is open.*

Proof: See for instance SPIVAK [Spi], page 39. □

2.40 Remark (**Local Coordinates for Submanifolds**)
Assume, for an open subset $W \subseteq \mathbb{R}^n$, that $0 \in F(W)$ is a regular value of $F \in C^1(W, \mathbb{R}^m)$. Then obviously $m \le n$, and by the implicit function theorem, we can find for $q \in M := F^{-1}(0)$ a neighborhood $U \subseteq W$ of q and a diffeomorphism $\varphi : U \to \mathbb{R}^n$ such that $\varphi(z)_i = 0$ for $n - m < i \le n$ and all $z \in U \cap M$.

The first $n - m$ components of φ, restricted to $U \cap M$, will serve as **local coordinates** of the submanifold M of $\mathbb{R}^n$. At q for instance, one can always use an appropriate choice of $n - m$ of the n cartesian coordinates. ◊

2.41 Example (**Sphere**)
Zero is a regular value of the function $F : \mathbb{R}^3 \to \mathbb{R}$, $F(x) := \|x\|^2 - 1$, because $\|\mathrm{D}F(x)\| = \|2x\| = 2$ for the pre-images $x \in F^{-1}(0) = S^2$. The north pole

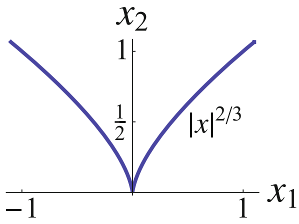
$q := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \in S^2$ has $U := \{x \in \mathbb{R}^3 \mid x_3 > 0\}$ as a neighborhood. The diffeomorphism $\varphi : U \to V,\ \varphi(x) := \big(x_1, x_2, F(x)\big)$ onto the image has the mapping $\varphi^{-1}(y) = \big(y_1, y_2, \sqrt{y_3 + 1 - y_1^2 - y_2^2}\big)$ as its inverse.

As we can make an analogous construction for each point $q \in S^2$ by an appropriate rotation, we have shown that S^2 is a 2-dimensional submanifold of $\mathbb{R}^3$. ♢

2.42 Example As a counterexample, take

$$F : \mathbb{R}^2 \to \mathbb{R}\ ,\ F(x_1, x_2) := x_1^2 - x_2^3,$$

so that $F^{-1}(0) = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 = x_2^3\}$. Here, too, F is infinitely differentiable, but 0 is not a regular value of F. The level set $F^{-1}(0)$ is shown on the right; it is not a submanifold. ♢



2.43 Definition *A continuous dynamical system $\Phi : G \times M \to M$ is called* **differentiable** *if M is a differentiable manifold and Φ is continuously differentiable.*

2.44 Remarks (Differentiable Dynamical Systems)

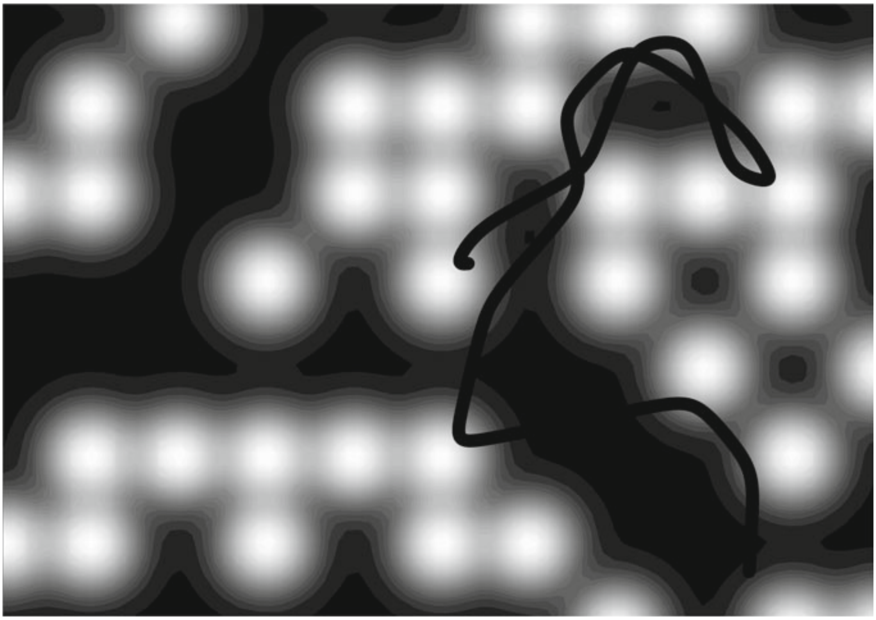
1. The methods of analysis can therefore be applied to differentiable dynamical systems. One example is the investigation of stability. For instance, Lyapunov's theorem (Theorem 7.6) concludes from the eigenvalues of the total derivative of a vector field at an equilibrium to the asymptotic stability with respect to the flow defined by this vector field.
2. The diffeomorphisms $f \in C^r(M, M)$ of a manifold M (see Def.A.36) form a group under composition, which is called the *diffeomorphism group* $\mathrm{Diff}^r(M)$. If M is compact, one can consider $\mathrm{Diff}^\infty(M)$ as an infinite dimensional Lie group whose Lie algebra is the space $\mathcal{X}(M)$ of smooth vector fields with Lie bracket (10.20).
   Thus a differentiable dynamical system $\Phi : G \times M \to M$ is a group homomorphism
   $$G \to \mathrm{Diff}(M) \quad , \quad g \mapsto \Phi_g.$$
   This point of view is sometimes helpful in understanding dynamical systems.
   - It makes sense to study which properties of dynamical systems are typical. For instance, it is true for compact M that those diffeomorphisms $F \in \mathrm{Diff}(M)$ that have only finitely many fixed points form an open and dense subset of $\mathrm{Diff}(M)$, in the topology of uniform C^r convergence.
     More generally, we call a property that can apply to discrete dynamical systems *generic* if the subset of $\mathrm{Diff}(M)$ defined by this property is the intersection of countably many open and dense subsets.
   - Instead of the point set topology of $\mathrm{Diff}(M)$, one can study its algebraic topology and for instance observe that the diffeomorphisms $F \in \mathrm{Diff}(S^1)$ of

the circle $S^1 \subset \mathbb{C}$ either lie in the connected component of the identity, or else of the conjugation map $S^1 \to S^1,\ z \mapsto \overline{z}$. $\Diamond$

2.45 Exercises (**Diffeomorphism Group**) Show that the diffeomorphism group $\mathrm{Diff}(M)$ of a connected manifold M operates transitively, i.e., for every $x, y \in M$, there exists some $f \in \mathrm{Diff}(M)$ for which $f(x) = y$.
Hint: First prove that for all y in a small neighborhood of $x \in M$, there exists a vector field on M whose time-one flow f is a diffeomorphism with $f(x) = y$. $\Diamond$

2.46 Literature An early survey article on differentiable dynamical systems that is useful reading is [Sm1] by STEVEN SMALE. $\Diamond$

Chapter 3
Ordinary Differential Equations



Motion in a random potential (see page 248)

Differential equations are as varied as the phenomena of nature described by them. This chapter begins by organizing different notions for differential equations and by transforming differential equations into a normal form (for explicit differential equations of first order). After this, existence, uniqueness, and smoothness of the solution to the initial value problem will be investigated. In doing so, explicit solution techniques are not in focus yet. Readers who have the corresponding prerequisites may skip Chapter 3.6 without problems.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_3

3.1 Definitions and Examples

We begin with (somewhat informal) definitions and a rough classification:

3.1 Definition

- *A* **differential equation (DE)** *is an equation in which derivatives of one or several functions of one or several variables occur. In a DE, the unknown quantities are the functions.*
- *If the functions depend on only* **one** *variable (often called* **time***), the differential equation is called* **ordinary (ODE)***, otherwise* **partial (PDE)**.
- *If several functions are to be found, we talk about a* **system of differential equations***, otherwise a* **single DE**.

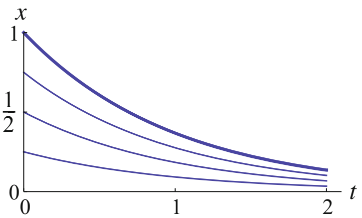
3.2 Examples (Differential Equations)

1. For instance, for $c > 0$, the single ordinary differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = -c\,x(t)$$

describes radioactive decay; here x is the amount of a substance as a function of time t, and c is the decay constant. If the amount at time $t = 0$ equals $x_0 \in \mathbb{R}$, then the unique solution is

$$x(t) = x_0\,e^{-ct} \qquad (t \in \mathbb{R}).$$



So we obtain a one-parameter family of solutions, which depends linearly on the initial value x_0 (see figure).

2. The orbit of an object that is thrown within the constant gravitation of the earth with acceleration of gravity[1] $g > 0$ is described, neglecting air resistance, by the system of ordinary differential equations

$$\frac{\mathrm{d}^2x_1}{\mathrm{d}t^2}(t) = 0 \quad , \quad \frac{\mathrm{d}^2x_2}{\mathrm{d}t^2}(t) = -g.$$

Here, x_1 denotes the horizontal component and x_2 the vertical component of the position as a function of time t.
For an initial position $x_0 = \binom{x_{1,0}}{x_{2,0}} \in \mathbb{R}^2$ and initial velocity $v_0 = \binom{v_{1,0}}{v_{2,0}} \in \mathbb{R}^2$, the *solution* is:

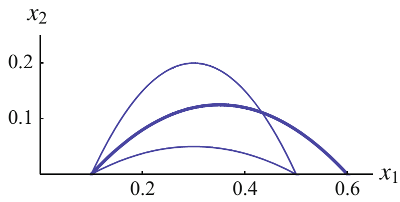
$$x_1(t) = x_{1,0} + v_{1,0}t \quad , \quad x_2(t) = x_{2,0} + v_{2,0}t - \tfrac{1}{2}gt^2 \qquad (t \in \mathbb{R}).$$

[1] At surface level, $g = 9.81\,\mathrm{m/s}^2$.

This corresponds to the velocities

$$v_1(t) := \tfrac{\mathrm{d}}{\mathrm{d}t}x_1(t) = v_{1,0} \quad , \quad v_2(t) := \tfrac{\mathrm{d}}{\mathrm{d}t}x_2(t) = v_{2,0} - gt \qquad (t \in \mathbb{R}).$$

The figure shows several trajectories with a common initial position and common absolute value of the initial velocity, but differing in the direction of the initial velocity. The trajectory at an initial angle $\alpha = \pi/4$ reaches farthest, because for time $t := 2v_{2,0}/g$, one has $x_2(t) = x_{2,0}$, and in view of $v_0 = \|v_0\| \left(\begin{smallmatrix}\cos\alpha \\ \sin\alpha\end{smallmatrix}\right)$, the horizontal distance is



$$x_1(t) - x_{1,0} = \frac{2v_{1,0}v_{2,0}}{g} = \frac{\|v_0\|^2 \sin(2\alpha)}{g}\,.$$

3. The one dimensional *wave equation* $\frac{\partial^2 u}{\partial t^2}(x,t) = c^2 \frac{\partial^2 u}{\partial x^2}(x,t)$, with the *wave speed* $c > 0$ as a parameter, is an example of a partial differential equation. For arbitrary functions $f_\pm \in C^2(\mathbb{R})$, the function

$$u(x,t) := f_+(x - ct) + f_-(x + ct) \qquad (x, t \in \mathbb{R})$$

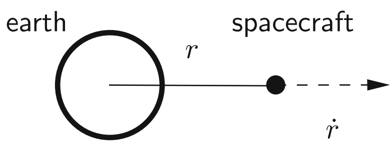
is a solution. An application in physics is the travel of electrical signals $f_\pm$ in a telegraph wire, where x denotes the position and t the time. ◊

In this book, we will only deal with *ordinary differential equations*, or *ODEs*.

3.3 Definition *The highest order of a derivative occuring in a DE will be called the* **order of the differential equation**.

3.4 Examples (Order of Differential Equations)

1. The ODE from Example 3.2.1 is of first order.
2. The one from 3.2.2 is of second order.
3. The differential equation (1.3), as well as its radial component, $\frac{\mathrm{d}^2 r}{\mathrm{d}t^2}(t) = -\frac{\gamma}{r^2(t)}$, which is obtained by specializing to the case of vanishing angular momentum, are of second order. The latter ODE describes, for example, the motion of a spacecraft that moves away radially with speed $\frac{\mathrm{d}r}{\mathrm{d}t}$ from the center of the earth. Then the parameter γ is the product of the constant of gravitation and the mass of the earth, $M > 0$; and r denotes the distance of the spacecraft from the center of the earth, and $\dot{r} = \frac{\mathrm{d}}{\mathrm{d}t}r$ the radial velocity. ◊



3.5 Definition

- *A system of ordinary differential equations for the functions* $x_1, \ldots, x_m$ *is called* **linear** *if it is of the form*

$$\sum_{i=0}^{n} A^{(i)}(t)\, x^{(i)}(t) \;=\; b(t)\,.$$

Here, $x^{(i)} := (\frac{\mathrm{d}^i}{\mathrm{d}t^i}x_1, \ldots, \frac{\mathrm{d}^i}{\mathrm{d}t^i}x_m)^\top$ *denotes the vector of the* i^{th} *derivatives;* $t \mapsto A^{(i)}(t) \in \mathrm{Mat}(m, \mathbb{R})$ *and* $t \mapsto b(t) \in \mathbb{R}^m$ *are given matrix or vector valued functions respectively.*
- *In all other cases, the system of ordinary differential equations is called* **nonlinear***.*
- *A linear ODE is called* **homogeneous** *if* $b(t) = 0$ *for all* t*, otherwise it is called* **inhomogeneous***.*
- *The components* b_l *of* b *are called* **forcings***, or* **source terms***.*

For instance, Example 3.2.1 is linear homogeneous, Example 3.2.2 is linear inhomogeneous, and Example 3.4.3 is nonlinear.

From now on, we will introduce many notions for single differential equations only. Most of them will carry over immediately to systems of ODEs.

3.6 Definition

1. *An ordinary differential equation is called* **implicit** *if it is given in the form*

$$F\left(t, x, x', \ldots, x^{(n)}\right) = 0\,, \tag{3.1.1}$$

and **explicit** *if it is given in the form*

$$x^{(n)} = f\left(t, x, x', \ldots, x^{(n-1)}\right)\,. \tag{3.1.2}$$

2. *An* n *times differentiable function defined on an open interval* I*, say* $x : I \to \mathbb{R}$*, is called an* **explicit solution** *of the ODE (3.1.1) or (3.1.2) respectively if one has*

$$F\big(t, x(t), x'(t), \ldots, x^{(n)}(t)\big) = 0 \qquad (t \in I)$$

$$\text{or} \quad x^{(n)}(t) = f\big(t, x(t), x'(t), \ldots, x^{(n-1)}(t)\big) \qquad (t \in I)$$

respectively.

Examples 3.2.1–2 and Example 3.4.3 are explicit differential equations. For Examples 3.2.1–2, we have also given the explicit solutions.

3.7 Remark (Notion of a Solution) We are familiar with algebraic equations like $ax^2 + bx + c = 0$, with given coefficients $a, b, c \in \mathbb{R}$; they are *blank statements* depending on variables x taken from the set $\mathbb{R}$, such that a statement (either true or false) arises, when we plug in a number $x \in \mathbb{R}$.

Similarly we can consider a differential equation, e.g., of type (3.1.2) with a continuous f to be a blank statement about variables from the set $C^n(I, \mathbb{R})$, where solutions are the ones that make the statements true, see WÜST [Wu], Chapter 5.2. ♢

3.8 Example (Implicit and Explicit Solutions)
$y\frac{\mathrm{d}y}{\mathrm{d}x} + x = 0$ is an example of an implicit nonlinear DE. The general solution is the equation of a circle $x^2 + y^2 = c \geq 0$, but it is given in *implicit* form. *Explicit* solutions are: $y(x) = \pm\sqrt{c - x^2}$ for $|x| < \sqrt{c}$, so

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mp\frac{x}{\sqrt{c - x^2}} = -\frac{x}{y}. \qquad \diamondsuit$$

The following definitions are somewhat heuristic:

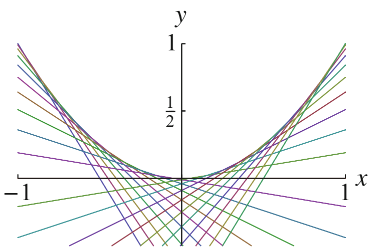
3.9 Definition

- *A single solution (without an arbitrary constant) is called a* **special** *or* **particular** *solution.*
- *A solution of a differential equation of the* n^{th} *order is called a* **general** *solution if it contains n arbitrary independent parameters;*
- *a solution dependent on parameters is called* **complete** *, if all particular solutions arise from it by choosing appropriate values for the parameters.*
- *A particular solution that does not belong to a parameter dependent solution is called* **singular**.

Examples 3.2.1 and 2: The general (also complete) solutions have been given. A particular solution to #2 is for instance $x_1(t) = 0,\ x_2(t) = -\frac{1}{2}gt^2$.

Note: In #2, there are *four* parameters $x_{1,0}, x_{2,0}, v_{1,0}, v_{2,0}$, because these are two differential equations of second order, $2 \times 2 = 4$.
Example 3.8.: Here, $c \geq 0$ is the parameter of the general (and complete) solution.

3.10 Example (Implicit Differential Equation)
$(y')^2 - 4xy' + 4y = 0$ is an implicit nonlinear ODE of first order.
The *general* solution $y(x) = 2cx - c^2,\ c \in \mathbb{R}$ is a family of straight lines, parametrized by c. But this is *not* the *complete* solution, because there still exists the *singular* solution $y(x) = x^2$. This parabola is the envelope of the family of straight lines, see the figure. ♢



Questions:
- How can one find solutions?
- How does one determine that all solutions have been found?

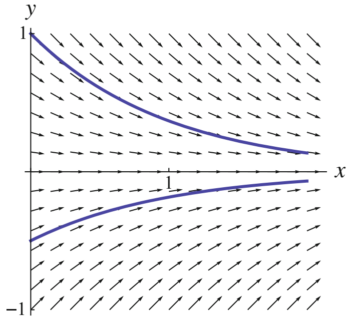
These questions have been studied by many mathematicians since Newton's days (and we will study them in this chapter).[2]

Let us first study single explicit differential equations of first order

$$y' = f(x, y) \qquad (x, y) \in U \subseteq \mathbb{R}^2,\ U \text{ open}.$$

Geometric Interpretation: If at each point $(x, y) \in U$ one draws a straight line segment with slope $f(x, y)$, then the graph

$$\text{graph}(\tilde{y}) = \{(x, \tilde{y}(x)) \mid x \in I\} \subset U$$



of every particular solution $\tilde{y} : I \to \mathbb{R}$ to the differential equation is the image of a curve $I \to U,\ x \mapsto (x, \tilde{y}(x))$ that is everywhere *tangential* to the local straight line segments.

Example: $y' = -cy$, so $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = -cy$, see the figure above.

In conclusion, to find the particular solution passing through the point (x, y), one moves, beginning at (x, y), in directions that are always tangential to the field of directions.

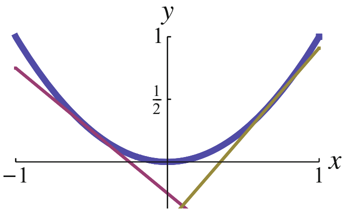
Caution: How do we know that there is only *one* solution curve passing through each point $(x, y) \in U \subset \mathbb{R}^2$?

3.11 Examples (Counterexamples to unique solvability)

1. **Implicit Differential Equation from Example 3.10**

$$(y')^2 - 4xy' + 4y = 0 \qquad (3.1.3)$$

In this example, there are two solution curves passing through each point (x_0, y_0) below the parabola $y = x^2$, namely straight lines tangential to the parabola. Their slopes correspond to the two solutions of the quadratic equation (3.1.3) for y' at the point (x_0, y_0). There are no solutions above the graph of the parabola.

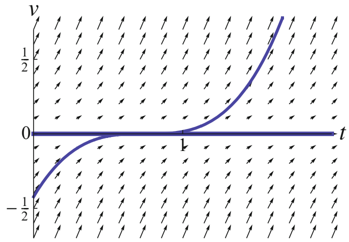


2. **Explicit Differential Equation with f not Lipschitz continuous**
   A general solution of the ODE $\dot{v} = f(v)$ with $f(v) := 3\sqrt[3]{v^2}$ is given by

$$v(t) = (t - c)^3,\ c \in \mathbb{R}.$$

[2]Numerical methods for the solution of differential equations are discussed, for instance, in GRIFFITHS and HIGHAM [GH].

But in addition, there is a singular solution, $v(t) = 0$.
Therefore there are at least two solution curves passing through each point on the t-axis!
It is noteworthy in this example that the function f, while being continuous, is not differentiable at 0.



Comparison with the (unique) case $f(t) := |t|$ suggests that it is not the lack of differentiability, but the lack of Lipschitz continuity, that causes the failure of uniqueness.
In physics, this example models, for instance, the growth of the volume v of raindrops due to condensation of water vapor on the surface. It is assumed here that the rate of condensation is proportional to the surface of the water drop, i.e., proportional to $v^{2/3}$. Therefore this model cannot explain anything about the actual *formation* of drops. ◊

3.12 Exercises (Single Differential Equations of First Order)

1. Sketch the graphs of the velocity functions $f_i : \mathbb{R} \to \mathbb{R}$ with

$$f_1(x) := (x^2 - 1)^2 \quad , \quad f_2(x) := (x^2 + 1)^2 .$$

   Determine the fixed points and the minimal invariant sets of the differential equations $\dot{x} = f_i(x)$.
   Without solving the differential equations explicitly, describe the qualitative behavior of their solutions $x_i(t, x_0)$ for times t and initial value x_0.
2. Determine, in dependence on $\alpha \geq 0$ and the initial value $x_0 > 0$, the maximal time interval on which the initial value problem $\dot{x} = f(x)$ for $f : \mathbb{R}^+ \to \mathbb{R}$, $f(x) := x^\alpha$ has a solution. ◊

After this informal survey of phenomena occurring in ordinary differential equations, we now show in mathematical rigor the (local) existence and uniqueness of solutions to sufficiently regular explicit ODEs of first order. We will later see that this also answers the same question for explicit ODEs of higher order.

3.2 Local Existence and Uniqueness of the Solution

We will now see that the differential equation $\dot{x} = f(t, x)$ is locally uniquely solvable if f has a bit more regularity than merely being continuous. We will study the n-dimensional situation right away:

3.13 Definition

- *If $U \subseteq \mathbb{R}_t \times \mathbb{R}^n_x$ is open and $f : U \to \mathbb{R}^n$ is continuous, then U is called* **extended phase space**, *f is called a* **time dependent vector field**, *and the equation*

$$\boxed{\dot{x} = f(t, x)}$$

is called a **nonautonomous** *or* **explicitly time dependent** *differential equation.*
- *If in particular $U = \mathbb{R}_t \times \tilde{U}$ with* **phase space** *$\tilde{U} \subseteq \mathbb{R}^n_x$ open, and if f has the form $f(t, x) = \tilde{f}(x)$, then we call the ODE* **autonomous** *or a* **dynamical system**.
- *A differentiable function $\varphi : I \to \mathbb{R}^n_x$ on the interval $I \subseteq \mathbb{R}_t$ is called a* **solution** *to the differential equation if* $\mathrm{graph}(\varphi) \subset U$ *and*

$$\left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{t=\tau} = f\big(\tau, \varphi(\tau)\big) \qquad (\tau \in I).$$

- *For $(t_0, x_0) \in U$ a map $\varphi : I \to \mathbb{R}^n_x$* **satisfies** *the* **initial condition** *(t_0, x_0) if $t_0 \in I$ and $\varphi(t_0) = x_0$. φ* **solves the initial value problem (IVP)** *if*

$$\left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{t=\tau} = f\big(\tau, \varphi(\tau)\big) \quad (\tau \in I) \quad \textit{and} \quad \varphi(t_0) = x_0 \,. \tag{3.2.1}$$

- *The time dependent vector field $f : U \to \mathbb{R}^n$ satisfies*
  - **a global Lipschitz condition** *with constant L if*

$$\|f(t, x_0) - f(t, x_1)\| \le L\,\|x_0 - x_1\| \qquad \big((t, x_i) \in U\big)\,,$$

  - *and* **a local Lipschitz condition** *if each point (τ, x) in U has a neighborhood $V \subseteq U$ such that*

$$\|f(t, x_0) - f(t, x_1)\| \le L\,\|x_0 - x_1\| \qquad \big((t, x_i) \in V\big) \tag{3.2.2}$$

  for some constant $L = L(\tau, x)$.

3.14 Lemma (Local Lipschitz Condition)
If the time dependent vector field $f : U \to \mathbb{R}^n_x$ is continuously differentiable, then a local Lipschitz condition (3.2.2) is satisfied on every compact and convex subset $V \subseteq U$ of the extended phase space, with the Lipschitz constant

$$L := \sup_{(t,x)\in V} \|\mathrm{D}_x f(t, x)\|\,.$$

Proof:
- Since $\mathrm{D}_x f : V \to \mathrm{Mat}(n, \mathbb{R})$ is continuous and V is compact, one gets $L < \infty$.
- For the points $x_s := (1 - s)x_0 + s x_1$ $(s \in [0, 1])$ on the segment, one obtains $(t, x_s) \in V$ from the assumption of convexity.

• The fundamental theorem of calculus gives us: $f(t,x_1) - f(t,x_0) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}s} f(t,x_s)\,\mathrm{d}s = \int_0^1 \mathrm{D}_x f(t,x_s)\frac{\mathrm{d}x_s}{\mathrm{d}s}\,\mathrm{d}s = \int_0^1 \mathrm{D}_x f(t,x_s)(x_1 - x_0)\,\mathrm{d}s$, which implies (3.2.2):
$\|f(t,x_1) - f(t,x_0)\| \le \int_0^1 \|\mathrm{D}_x f(t,x_s)\|\,\mathrm{d}s\ \|x_1 - x_0\| \le L\|x_1 - x_0\|$. □

3.15 Remarks (Existence and Uniqueness)

1. Observe that in (3.2.2), Lipschitz continuity is only required with respect to the variable x.
2. By the Picard-Lindelöf theorem (Theorem 3.17), the local Lipschitz condition is already sufficient for existence and uniqueness (see Definition 3.16) of a solution (local in time) to the initial value problem.
   The mere *existence* of such a solution already follows from our general hypothesis that the time dependent vector field f is continuous (Peano's theorem).
3. As has already been observed, a solution $\varphi : I \to \mathbb{R}^n_x$ to the IVP gives rise to a solution $\varphi\restriction_{\tilde{I}}$ by restriction to a smaller interval $\tilde{I} \subseteq I$ that still contains t_0. Strictly speaking, this is a different solution because the domains of the two functions φ and $\varphi\restriction_{\tilde{I}}$ are different. In this sense, of course, the solution to the initial value problem is not unique.
   However, as we are looking for the largest time interval I on which a solution to (3.2.1) can be defined (see Sect. 3.5), we are not interested in this trivial distinction between solutions and thus remove it by definition: ◊

3.16 Definition *We say the solution to the initial value problem (3.2.1) is* **unique** *if any two solutions $\varphi_1 : I_1 \to \mathbb{R}^n_x$ and $\varphi_2 : I_2 \to \mathbb{R}^n_x$ to the initial value problem coincide on the interval $I_3 := I_1 \cap I_2$, i.e., satisfy the condition*

$$\varphi_1\restriction_{I_3} = \varphi_2\restriction_{I_3} .$$

In the proof of Theorem 3.17, we will find the unique local solutions to the initial value problem as fixed points of a contraction mapping on some space of continuous functions. In order to guarantee the existence of the fixed point, we use Banach's fixed point theorem (Theorem D.3 in the appendix) and accordingly the completeness of said space of continuous functions, which is stated in Theorem D.1 on page 543.

3.17 Theorem (Picard-Lindelöf)
Assume that the time dependent vector field $f : U \to \mathbb{R}^n$ on the extended phase space $U \subseteq \mathbb{R}_t \times \mathbb{R}^n_x$ satisfies a Lipschitz condition on U.
Then for $(t_0, x_0) \in U$, there exists an $\varepsilon > 0$ such that the initial value problem

$$\dot{x} = f(t,x) \quad , \quad x(t_0) = x_0 \tag{3.2.3}$$

has a unique solution $\varphi : [t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}^n_x$.

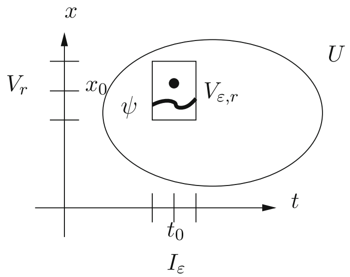
Proof: Let us denote the yet to be determined time interval as $I \equiv I_\varepsilon := [t_0 - \varepsilon, t_0 + \varepsilon]$.

- If such a solution exists, it has to satisfy the integral equation

$$\varphi(t) = x_0 + \int_{t_0}^{t} f\big(s, \varphi(s)\big)\,\mathrm{d}s \qquad (t \in I), \tag{3.2.4}$$

as can be seen by taking a definite integral and plugging in t_0 respectively.
On the other hand, by the fundamental theorem of calculus, every continuous solution to (3.2.4) is automatically differentiable and therefore a solution to the initial value problem (3.2.3).

- Now the solution φ is to be found as a fixed point of a certain mapping A. A is to map continuous curves in phase space to such curves again. In order to choose the domain of A conveniently, let the region in phase space be the closed ball $V \equiv V_r := \overline{U_r(x_0)}$. We define



$$V_{\varepsilon,r} := I_\varepsilon \times V_r$$

and choose r small enough for $V_{r,r}$ to be contained in U.

Moreover, let $L > 0$ be a Lipschitz constant of $f\restriction_{V_{r,r}}$,

$$N := \max_{(t,x)\in V_{r,r}} \|f(t,x)\| \quad \text{and} \quad \varepsilon := \min\left(r, \frac{r}{N}, \frac{1}{2L}\right). \tag{3.2.5}$$

Then[3] in particular, $V_{\varepsilon,r} \subseteq V_{r,r} \subseteq U$. Let

$$M := C(I, V) \subseteq C(I, \mathbb{R}^n)$$

again denote the metric space of continuous functions $\psi : I \to V$, equipped with the sup metric

$$d(\psi, \varphi) := \sup_{t\in I} \|\psi(t) - \varphi(t)\|\,.$$

By Theorem D.1, (M, d) is a complete metric space (since $V \subseteq \mathbb{R}^n$ is closed).

- We define by

$$\boxed{(A\psi)(t) := x_0 + \int_{t_0}^{t} f\big(s, \psi(s)\big)\,\mathrm{d}s \qquad (t \in I)}$$

a mapping $A : M \to C(I, \mathbb{R}^n)$. We first show that its image is again contained in M. The distance between $(A\psi)(t)$ and x_0 is

[3] With the stipulation $r/N := +\infty$ when $N = 0$.

$$\left\| \int_{t_0}^{t} f\big(s, \psi(s)\big)\,\mathrm{d}s \right\| \le \left| \int_{t_0}^{t} \big\| f\big(s, \psi(s)\big) \big\| \,\mathrm{d}s \right|$$
$$\le \left| \int_{t_0}^{t} \max_{(\tilde{t},x)\in V_{r,r}} \| f(\tilde{t}, x)\| \,\mathrm{d}s \right| \le |t - t_0| N \le \varepsilon N \le r\,,$$

where we have used the definition (3.2.5) of ε.
So one has $(A\psi)(t) \in \overline{U_r(x_0)} = V$ for all times $t \in I$, hence $A\psi \in M$.

- Now the missing hypothesis in Banach's fixed point theorem is that this so called *Picard map*
$$A : M \to M$$
is a contraction, i.e., for an appropriate $0 < \theta < 1$, it satisfies
$$d\big(A\varphi, A\psi\big) \le \theta\, d(\varphi, \psi) \qquad (\varphi, \psi \in M).$$
Indeed, we conclude
$$\begin{aligned} &d\big(A\varphi, A\psi\big) = \sup_{t\in I} \|A\varphi(t) - A\psi(t)\| \\ &= \sup_{t\in I} \left\| \int_{t_0}^{t} \Big[f\big(s, \varphi(s)\big) - f\big(s, \psi(s)\big)\Big] \,\mathrm{d}s \right\| \le \varepsilon \sup_{s\in I} \big\| f\big(s, \varphi(s)\big) - f\big(s, \psi(s)\big)\big\| \\ &\le \varepsilon\, L \sup_{s\in I} \|\varphi(s) - \psi(s)\| = \varepsilon\, L\, d(\varphi, \psi) \le \tfrac{1}{2}\, d(\varphi, \psi)\,. \end{aligned}$$
In the last inequality, we have again used the definition (3.2.5) of ε.
In conclusion, A is a contraction mapping on the complete metric space M.
- By Banach's fixed point theorem (see page 544), A has a unique fixed point $\varphi \in M$. This function φ satisfies the integral equation (3.2.4) and hence solves the initial value problem. □

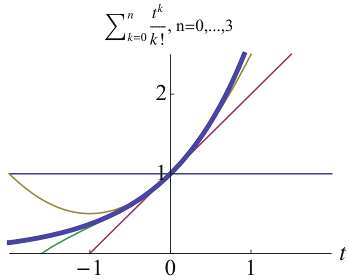
The *Picard iteration* $x_{i+1} := Ax_i$ used here to find the fixed point $\varphi = \lim_{i\to\infty} x_i$ of A can also be used to solve differential equations:

3.18 Examples (Picard Iteration)

1. We approximate the solution to the initial value problem
$\dot{x} = x$, $x(0) = x_0 \in \mathbb{R}$ by
$$x_0(t) := x_0 \quad \text{and} \quad x_{i+1}(t) := x_0 + \int_0^t x_i(s)\,\mathrm{d}s \qquad (t \in \mathbb{R}),$$

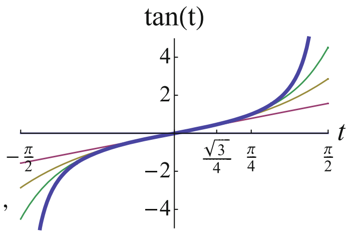
i.e., (see figure)

$$x_1(t) = x_0\ (1+t)$$
$$x_2(t) = x_0\ \left(1+t+\frac{t^2}{2}\right)$$
$$\vdots$$
$$x_n(t) = x_0\ \sum_{i=0}^{n} \frac{t^i}{i!}\,.$$



As $x(t) = x_0 \sum_{i=0}^{\infty} \frac{t^i}{i!} = x_0 \cdot e^t$, for all $t \in \mathbb{R}$, the n^{th} iterate $x_n(t)$ converges to the solution $x(t)$, and it does so uniformly on every compact time interval (but not uniformly on all of $\mathbb{R}$).

2. The initial value problem $\dot{x} = 1 + x^2$, $x_0 = 0$ has a solution only on the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, namely the tangent function.
We optimize the constants in the proof of the Picard-Lindelöf theorem: For $r > 0$, we have $N = \max_{|x|\le r} \|f(x)\| = 1 + r^2$, and the Lipschitz constant $L = \max_{|x|\le r} \|f'(x)\| = 2r$. So by definition (3.2.5), we have $\varepsilon = \min\left(\frac{r}{1+r^2}, \frac{1}{4r}\right)$. ε will be maximal for $r = \frac{1}{\sqrt{3}}$, i.e., $\varepsilon = \frac{\sqrt{3}}{4}$. For times $|t| < \frac{\sqrt{3}}{4}$, we can therefore guarantee convergence. The Picard iteration with initial value $x_0(t) := x_0$ yields



$$\begin{aligned}
x_0(t) &= 0,\\
x_1(t) &= (Ax_0)(t) = \textstyle\int_0^t \left[1 + x_0^2(s)\right] ds = t,\\
x_2(t) &= (Ax_1)(t) = \textstyle\int_0^t \left[1 + s^2\right] ds = t + t^3/3,\\
x_3(t) &= (Ax_2)(t) = \textstyle\int_0^t \left[1 + (s + s^3/3)^2\right] ds\\
&= t + \tfrac{t^3}{3} + \tfrac{2}{15}t^5 + \tfrac{1}{63}t^7,
\end{aligned}$$

etc. As a matter of fact, this sequence of functions will converge to the tangent function not only on the interval $[-\varepsilon, \varepsilon]$, but on the entire interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. ◊

3.19 Exercise (Picard-Lindelöf) We consider the differential equation

$$\dot{x} = f(x) \quad \text{with} \quad f: \mathbb{R} \to \mathbb{R}\,,\ f(x) := \exp(-x)$$

and initial condition $x(0) = 0$. As f is locally Lipschitz continuous, there exists, by Theorem 3.17, some $\varepsilon > 0$ and a function $\varphi : [-\varepsilon, \varepsilon] \to \mathbb{R}$ satisfying the initial value problem.

(a) Find a lower bound for ε guaranteed by this theorem.
(b) What does the Picard iteration look like for this initial value problem?
(c) What is the maximal solution to the initial value problem? ◊

In applications of differential equations to science and technology, we usually do not know the initial values precisely. The following theorem tells us that this is not actually necessary.

3.20 Theorem *Under the assumptions of Theorem 3.17 (Picard-Lindelöf), each point $(T_0, X_0) \in U$ of the extended phase space has a compact neighborhood $V \subset U$ and an interval $I_\varepsilon := [-\varepsilon, \varepsilon]$ such that the family*

$$\Phi : I_\varepsilon \times V \to U \quad , \quad (s; t_0, x_0) \mapsto \varphi(t_0 + s)$$

of solutions to the initial value problem (3.2.3) is a continuous mapping. Therefore, the solutions depend continuously on their initial values and initial time.

Proof:

- For small $R > 0$ and $\varepsilon > 0$, the set $[T_0 - 2\varepsilon, T_0 + 2\varepsilon] \times \overline{U_R(X_0)}$ is a subset of the extended phase space U.
  $V_{\varepsilon,r} := [T_0 - \varepsilon, T_0 + \varepsilon] \times \overline{U_r(X_0)}$, the set of initial values (t_0, x_0), is contained in this subset when $r \in (0, R)$. Moreover, the set is bounded and closed, hence compact. Now, instead of using the space of curves for the Picard iteration, we use the metric space
  $$M := C\big(I_\varepsilon \times V_{\varepsilon,r} \, , \, \overline{U_R(X_0)}\big)$$
  with the sup metric
  $$d(\Phi, \Psi) := \sup\{\|\Phi(t; y) - \Psi(t; y)\| \mid (t; y) \in I_\varepsilon \times V_{\varepsilon,r}\} \qquad (\Phi, \Psi \in M).$$

- (M, d) is a complete metric space, because:

 (a) The target space is a closed subset of $\mathbb{R}^n$, hence complete.
 (b) The domain $I_\varepsilon \times V_{\varepsilon,r}$ is compact. Therefore $d(\Phi, \Psi) < \infty$, and each Cauchy sequence $(\Phi_m)_{m\in\mathbb{N}}$ in M has a *pointwise* limit
 $$\Phi : I_\varepsilon \times V_{\varepsilon,r} \to \overline{U_R(X_0)} \quad , \quad \Phi(t; y) := \lim_{m\to\infty} \Phi_m(t; y) \, .$$
 (c) Moreover, the $\varepsilon/3$ argument from the proof of Theorem D.1 carries over to this situation, so this limit Φ is continuous and thus $\Phi \in M$.

- For $\Psi \in M$, we consider the Picard mapping
  $$(A\Psi)(s; t_0, x_0) := x_0 + \int_0^s f\big(t_0 + \tau, \Psi(\tau; t_0, x_0)\big)\, \mathrm{d}\tau \quad \big((s; t_0, x_0) \in I_\varepsilon \times V_{\varepsilon,r}\big).$$

If Φ is a fixed point of A, then this means that

$$\Phi(0; t_0, x_0) = A\Phi(0; t_0, x_0) = x_0$$

and

$$\frac{\mathrm{d}}{\mathrm{d}s}\Phi(s; t_0, x_0) = \frac{\mathrm{d}}{\mathrm{d}s}(A\Phi)(s; t_0, x_0) = f\big(t_0 + s, \Phi(s; t_0, x_0)\big).$$

So the mapping $t \mapsto \Phi(t - t_0; t_0, x_0)$ solves the IVP with initial value (t_0, x_0).

- By the same reasoning as in the proof of Picard-Lindelöf, the Picard mapping will be a contraction

$$A : M \to M$$

for small parameters $\varepsilon, r > 0$. So it has a unique fixed point $\Phi \in M$ by Banach's fixed point theorem. □

3.3 Global Existence and Uniqueness of the Solution

It is plausible to study also differential equations on manifolds.

3.21 Definition

- *If $f : M \to TM$ is a (time independent) vector field on the manifold M (see Definition A.39), a curve $\varphi \in C^1(I, M)$ is called* **solution to the differential equation** $\dot{x} = f(x)$ *if* $\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t) = f\big(\varphi(t)\big)$ *for all times* $t \in I$.
- *A vector field $f : M \to TM$ on the manifold M is called* **complete** *if for all $x_0 \in M$, the initial value problem $\dot{x} = f(x),\ x(0) = x_0$ has a unique solution $\varphi : \mathbb{R} \to M$.*

In this section, we provide criteria for the completeness of vector fields. We begin with the case where the phase space is $M = \mathbb{R}^n$.

3.22 Examples

1. In Example 3.18.2 with the ODE $\dot{x} = f(x) = 1 + x^2$ for the tangent function, we see that the solution does not exist for all times, but rather diverges to infinity (beginning at 0) within time $\frac{\pi}{2}$, due to the fact that $x \mapsto f(x)$ exhibits superlinear growth. This is not a contradiction to the *local* Lipschitz continuity of f.
2. In contrast, in Example 3.18.1, $f(x) = x$ is even *globally* Lipschitz continuous, and this linear vector field is complete. This is true in full generality: ◊

3.23 Theorem

- *Lipschitz continuous vector fields $f : \mathbb{R}^n \to \mathbb{R}^n$ are complete.*
- *More generally, let $I \subseteq \mathbb{R}$ be an interval, and let the time dependent vector field $f : I \times \mathbb{R}^n \to \mathbb{R}^n$ satisfy the time dependent global Lipschitz condition*

$$\|f(t,x_1)-f(t,x_2)\| \le L(t)\|x_1-x_2\| \qquad \big(t\in I,\ x_1,x_2\in\mathbb{R}^n\big),$$

with $L: I \to \mathbb{R}^+$ *continuous. Then the initial value problem has a unique solution* $\varphi: I \to \mathbb{R}^n$ *for all initial values* $(t_0, x_0) \in I \times \mathbb{R}^n$.

Proof:
• The time independent case arises from the time dependent one by taking $I = \mathbb{R}$ and having L constant.
• It suffices to consider compact time intervals I, because each interval $I \ni t_0$ can be represented as a union of compact intervals $I_k \ni t_0$.
• Then as $L : I \to \mathbb{R}^+$ is continuous, it follows by means of the compactness hypothesis on the interval T that $\sup_{t\in I} L(t) < \infty$. So there exists a Lipschitz constant $\tilde{L} \ge 1$ with

$$\|f(t,x_1)-f(t,x_2)\| \le \tilde{L}\,\|x_1-x_2\| \qquad (t\in I,\ x_1,x_2\in\mathbb{R}^n).$$

We choose the radius $r \equiv r(x_0) := \sup_{t\in I}\|f(t,x_0)\| + \frac{1}{2\tilde{L}} < \infty$ of a ball around x_0 in such a way that inside this ball, the maximal velocity

$$N(x_0) := \max_{(t,x)\in I\times\overline{U_r(x_0)}}\ \|f(t,x)\|$$

satisfies the inequality

$$N(x_0) \le \max_{t\in I}\Big(\|f(t,x_0)\| + \max_{x\in\overline{U_r(x_0)}}\|f(t,x)-f(t,x_0)\|\Big) \le r(x_0)(1+\tilde{L}).$$

Then the constant ε from (3.2.5), which defines the time interval, is independent of x_0; however, it is no longer necessarily true that $[t_0-\varepsilon, t_0+\varepsilon] \subseteq I$:

$$\varepsilon(x_0) = \min\left(r(x_0)\,,\ \frac{r(x_0)}{N(x_0)}\,,\ \frac{1}{2\tilde{L}}\right) = \min\left(\frac{1}{2\tilde{L}}\,,\ \frac{1}{1+\tilde{L}}\,,\ \frac{1}{2\tilde{L}}\right) = \frac{1}{2\tilde{L}}\,.$$

• For arbitrary times $t_k \in I$ and initial values $x_k \in \mathbb{R}^n$, we can find, by Theorem 3.17, the unique local solution $\varphi_k : I_k \to \mathbb{R}^n$ to the initial value problem $\dot{x} = f(t,x)$ with $\varphi_k(t_k) := x_k$ on the interval $I_k := [t_k-\varepsilon, t_k+\varepsilon] \cap I$.

With the times $t_k := t_0 + \frac{\varepsilon}{2}k \in I$, we now define for $k \in \mathbb{N}$ the initial value $x_k := \varphi_{k-1}(t_k)$ in terms of the known solution φ_{k-1}. Analogously, we choose for integers $k < 0$ the initial condition $x_k := \varphi_{k+1}(t_k)$.
• So we have $\varphi_{k-1}(t_k) = \varphi_k(t_k)$. By uniqueness of the solution to the initial value problem in the sense of Definition 3.16, we obtain with

$$\varphi : I \to \mathbb{R}^n \quad , \quad \varphi(t) := \varphi_k(t) \text{ when } t \in I_k$$

a unique solution of the initial value problem $\dot{x} = f(t,x),\ x(t_0) = x_0$. □

3.24 Remarks

1. In analogy to Lemma 3.14, the following sufficient condition for global Lipschitz continuity is available: If the vector field satisfies $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, then f is Lipschitz continuous if and only if

$$\sup_{x \in \mathbb{R}^n} \|\mathrm{D}f(x)\| < \infty .$$

2. In particular, it follows that for all *linear differential equations*, the initial value problem

$$\dot{x} = Ax \quad , \quad x(0) = x_0$$

has a unique solution for all times, because the vector field $f(x) = Ax$ is Lipschitz continuous with constant $L = \|A\| := \sup_{v \in S^{n-1}} \|Av\|$, i.e., the matrix norm of $A \in \mathrm{Mat}(n, \mathbb{R})$. In Section 4.1, we show that this solution has the form $x(t) = \exp(At)x_0$.

3. Theorem 3.23 guarantees the unique global solvability, provided a time dependent Lipschitz constant $L : I \to \mathbb{R}^+$ exists. For linear inhomogeneous differential equations $\dot{x}(t) = A(t)x(t) + b(t)$, it even suffices to assume that $\|A\|$ and $\|b\|$ are locally integrable, see WEIDMANN [Weid], Theorem 2.1. ◊

3.25 Exercise Find the solution to the initial value problem $\dot{x} = \sin x$ with $x(0) = \pi/2$. Calculate $\lim_{t \to -\infty} x(t)$ and $\lim_{t \to \infty} x(t)$. ◊

The following example shows that Lipschitz continuity of the vector field on the entire phase space $\mathbb{R}^n$ is sufficient, but definitely not necessary for the vector field to be complete.

3.26 Example (Complete Vector Field)
The vector field $f : P \to \mathbb{R}^2,\ x \mapsto \|x\|^2 \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix}$ on the phase space $P = \mathbb{R}^2$ is smooth. However, it is not globally Lipschitz continuous, because its derivative

$$\mathrm{D}f : P \to \mathrm{Mat}(2, \mathbb{R}) \quad , \quad \mathrm{D}f(x) = \begin{pmatrix} 2x_1x_2 & x_1^2+3x_2^2 \\ -3x_1^2-x_2^2 & -2x_1x_2 \end{pmatrix}$$

is not bounded. Therefore, we cannot use Theorem 3.23 to prove its completeness. However, we observe that f is tangential to the circles $S^1_r := \{x \in P \mid \|x\| = r\}$ of radius r, because it is orthogonal to their normals: $\langle f(x), x \rangle = 0$.

In polar coordinates, $x_1 = r \cos \varphi$, $x_2 = r \sin \varphi$, the ODE $\dot{x} = f(x)$ becomes

$$\dot{r} = 0 \quad , \quad \dot{\varphi} = -r^2 .$$

The unique solution $r(t) = r_0$, $\varphi(t) = \varphi_0 - r_0^2\, t$ to the initial value problem exists for all times $t \in \mathbb{R}$. ◊

In this example, the decisive reason for global solvability was the possibility to restrict the vector field f to the *compact* submanifolds S^1_r of the phase space.

3.27 Theorem (Completeness)
Lipschitz continuous vector fields on compact manifolds are complete.

Proof:
• First note that *local* Lipschitz continuity (and this is what matters in view of the compactness hypothesis) is defined independent of coordinate charts. This is because changes of coordinate charts are diffeomorphisms of open subsets of $\mathbb{R}^n$, and their derivatives are bounded on compact sets.
• For each point $x \in M$, there is now a compact neighborhood K_x and an $\varepsilon_x > 0$, according to Theorem 3.20, such that the initial value problem has a unique solution on time interval $(-\varepsilon_x, \varepsilon_x)$ for each $x_0 \in K_x$. Also, in the maximal atlas for M, there are coordinate charts (U_x, φ_x) for each $x \in M$, with open domains $U_x \subset K_x$ containing x and mappings $\varphi_x : U_x \to \mathbb{R}^n$.

As the manifold M is compact, only finitely many of these charts are needed for an atlas (since every open cover of M has a finite subcover due to compactness). So we may assume that the index set I of the atlas $\{(U_i, \varphi_i) \mid i \in I\}$ of M is finite.
• As the minimum of the ε_i $(i \in I)$ is still positive, on can construct, with the piece-by-piece method from the proof of Theorem 3.23, a unique solution to the initial value problem with time interval $\mathbb{R}$. □

3.28 Exercise (Existence of the Flow) Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ be a smooth function such that $H^{-1}\big((-\infty, E]\big)$ is compact for all $E \in \mathbb{R}$.

(a) Give an example of such a function H.
(b) Show that the flow generated by the (Hamiltonian) differential equations

$$\dot{x}_j = -\frac{\partial H}{\partial x_{j+n}}(x) \quad , \quad \dot{x}_{j+n} = \frac{\partial H}{\partial x_j}(x) \qquad (j \in \{1, \dots, n\})$$

exists for all times. ◊

3.4 Transformation into a Dynamical System

Reduction to First Order

Differential equations of higher than first order can also be treated with the methods described, by converting an explicit ordinary differential equation of n^{th} order into a system of n ODEs of first order.

3.29 Theorem *The differential equation of order $n > 1$,*

$$\tfrac{\mathrm{d}^n x}{\mathrm{d}t^n} = F\left(t, x, \tfrac{\mathrm{d}x}{\mathrm{d}t}, \dots, \tfrac{\mathrm{d}^{n-1}x}{\mathrm{d}t^{n-1}}\right) \tag{3.4.1}$$

with $F \in C^1(\mathbb{R}^{n+1})$, is equivalent to the system of differential equations

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t,y) \quad \textit{with} \quad f(t,y) := \begin{pmatrix} y_2 \\ \vdots \\ y_n \\ F(t,y_1,\ldots,y_n) \end{pmatrix} \tag{3.4.2}$$

in the following sense:

- *If $\varphi : I \to \mathbb{R}$ solves (3.4.1), then $\psi := \begin{pmatrix} \varphi \\ \varphi' \\ \vdots \\ \varphi^{(n-1)} \end{pmatrix} : I \to \mathbb{R}^n$ solves (3.4.2).*
- *Conversely, if $\psi = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_n \end{pmatrix}$ solves (3.4.2), then ψ_1 solves (3.4.1).*

Proof: By definition, φ is n times differentiable, so $\psi : I \to \mathbb{R}^n$ is differentiable, and $\psi_k = \frac{\mathrm{d}}{\mathrm{d}t}\psi_{k-1} \quad (k = 2, \ldots, n)$,

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\psi_n = \tfrac{\mathrm{d}^n\varphi}{\mathrm{d}t^n} = F(t,\varphi,\varphi',\ldots,\varphi^{(n-1)}) = F(t,\psi_1,\ldots,\psi_n)\,.$$

The reasoning can be reversed. □

3.30 Exercise
Consider the one dimensional motion $\ddot{x} = x$. Find the solution with initial condition $(x_0, \dot{x}_0) = (1, -1)$. How much time does it take to reach $x = 0$? ♢

3.31 Example (Kepler Problem)
The 2-body problem from celestial mechanics describes, for instance, the motion of earth and sun around their common center of mass. It can be reduced to the *one-center problem* (mentioned in the introduction), which describes the motion of a point mass at location $x \in \mathbb{R}^3 \backslash \{0\}$ in the gravitation field of a celestial body that is located in the origin of the coordinate system. According to (1.3), one has

$$\ddot{x} = -\gamma \frac{x}{\|x\|^3}\,.$$

This system of ODEs of second order can be transformed into a differential equation of first order with the phase space $U := (\mathbb{R}^3\backslash\{0\}) \times \mathbb{R}^3$:

$$\dot{z} = f(z) := \begin{pmatrix} -\gamma \frac{v_x}{\|x\|^3} \end{pmatrix} \qquad \big(z = (x,v) \in U\big). \tag{3.4.3}$$

The phase space U is an open subset of $\mathbb{R}^6$, with a vector field $f \in C^\infty(U, \mathbb{R}^6)$.

This way, we can solve (3.4.3) by Picard iteration or some other method, for short times. The analogous statement applies to the n body problem (1.8). ♢

Transition to a Time Independent System
We can also reduce explicitly time dependent differential equations to autonomous ones. Instead of the ODE $\dot{x} = f(t,x)$ with $f : U \to \mathbb{R}^n$, $U \subseteq \mathbb{R}_t \times \mathbb{R}^n_x$ open, we consider the autonomous system of differential equations

$$\dot{y} = g(y) \quad \text{with} \quad g : U \to \mathbb{R}^{n+1} \quad , \quad y := \binom{s}{x} \qquad g(y) := \binom{1}{f(y)} . \tag{3.4.4}$$

In doing so, we increase the dimension of the phase space by one, by joining the time parameter s to the phase space point x. In full detail, Equation (3.4.4) is of the form

$$\tfrac{\mathrm{d}}{\mathrm{d}t} s = 1 \quad , \quad \tfrac{\mathrm{d}}{\mathrm{d}t} x = f(s, x) . \tag{3.4.5}$$

Now if $\psi : I \to U$ solves (3.4.4), then it follows with $\psi(t) = \binom{s(t)}{x(t)}$ that $s(t) = s(0) + t$; so the phase space coordinate s coincides with the time t up to an additive constant. The solution $\psi = \binom{s}{\tilde{x}}$ to the initial value problem $\dot{y} = g(y),\ \psi(0) = \binom{t_0}{x_0}$ will therefore yield a solution x to the initial value problem $\dot{x} = f(t, x),\ x(t_0) = x_0$. Simply let $x(t) := \tilde{x}(t - t_0)$.

Conversely, one can construct, from a solution to the IVP $\dot{x} = f(t, x), x(t_0) = x_0$, a solution to (3.4.4) by complementing it.

If $f : U \to \mathbb{R}^n$ is Lipschitz continuous (in all its arguments), then so is g. So the theorem about existence and uniqueness carries over. Note however that we do not require Lipschitz continuity with respect to time t in the case of nonautonomous ODEs.

Key Notions for Autonomous Differential Equations

If there exists, for every $x_0 \in M$, a unique solution $\varphi_{x_0} : \mathbb{R} \to M$ to the initial value problem $\dot{x} = f(x),\ x(0) = x_0$ on the manifold M, then the mapping

$$\Phi : \mathbb{R} \times M \to M \quad , \quad (t, x) \mapsto \varphi_x(t) \tag{3.4.6}$$

is called *phase flow*, or short *flow* of the differential equation. Then the flow is a continuous dynamical system in the sense of Definition 2.7.

However, it is often convenient to extend the use of notions that were so far defined only for dynamical systems to the situation when the solution to the initial value problem does not exist for all times:

3.32 Definition *Let the vector field $f : M \to TM$ on a manifold M be locally Lipschitz continuous. We consider the differential equation $\dot{x} = f(x)$.*

1. *Let $\varphi : I \to M$ be a solution curve of the differential equation with maximal time interval I. Then the image $\varphi(I) \subseteq M$ will be called* **orbit**. *For $x \in \varphi(I)$ the set $\mathcal{O}(x) := \varphi(I)$ is called the* **orbit through** *x.*
2. *$x_s \in M$ is called a* **singular point** *of the vector field f, if $f(x_s) = 0$.*
   If $x_s \in M$ is a singular point of f, then x_s is also called a **rest point** *or* **equilibrium** *of the differential equation.*
3. *$x \in M$ is called a* **periodic point** *with* **(minimal) period** *$T > 0$, if $\varphi_x(T) = x$ (and $\varphi_x(t) \neq x$ for $t \in (0, T)$).*
   An orbit $\mathcal{O}(x)$ is **closed**, *if $x \in M$ is a periodic point.*

3.33 Remarks

1. For a singular point x_s of the vector field f, the constant function $x(t) = x_s$ is the only solution to the initial value problem.
2. A point x_s is called *singular* not for the vector field having a singularity at that point, but rather because the *direction field*[4] $x \mapsto f(x)/\|f(x)\|$ is undefined there and can also in general not be extended continuously into x_s.
3. As we have already used in Theorem 3.27, the notion of *local* Lipschitz continuity, which is the basis of Definition 3.32, is well-defined also for vector fields on manifolds. ◊

3.34 Theorem *If the initial value problem $\dot{x} = f(x)$ defined by a locally Lipschitz continuous vector field $f : M \to TM$ has the solution $\Phi : \mathbb{R} \times M \to M$, then Φ is a continuous dynamical system (in the sense of Definition 2.16).*

Proof: According to Theorem 3.20, the mapping Φ given by (3.4.6) is continuous. The condition $\Phi_0 = \mathrm{Id}_M$ is satisfied due to the condition $\varphi_{x_0}(0) = x_0$ for a solution to the initial value problem. The composition property $\Phi_{t_1} \circ \Phi_{t_2} = \Phi_{t_1+t_2}$ $(t_1, t_2 \in \mathbb{R})$ follows from the uniqueness and translation invariance of the solution. □

3.35 Examples

1. For a real polynomial $f(x) = \prod_{i=1}^{n}(x - a_i)$ with zeros $a_1 < \ldots < a_n$, the differential equation $\dot{x} = f(x)$ has unique local solutions, and for $x \in [a_1, a_n]$ it even has solutions $t \mapsto \Phi_t(x)$ defined for all $t \in \mathbb{R}$.
   The equilibria are the points $a_1, \ldots, a_n$. For $x \in (a_i, a_{i+1})$ and $n - i$ even, we have $f\restriction_{(a_i,a_{i+1})} > 0$. In this case, one has therefore $\omega(x) = \{a_{i+1}\}$ and $\alpha(x) = \{a_i\}$. On the other hand, if $n - i$ is odd, hence $f\restriction_{(a_i,a_{i+1})} < 0$, one has conversely $\omega(x) = \{a_i\}$, $\alpha(x) = \{a_{i+1}\}$. There are no periodic points in this dynamical system.
2. For the system of differential equations (in polar coordinates (r, φ))
$$\dot{r} = r(1 - r^2) \quad , \quad \dot{\varphi} = 1 \, ,$$
   with phase space $\mathbb{R}^2$, the origin $(r = 0)$ is the only equilibrium, and $\{1\} \times [0, 2\pi)$ is the only periodic orbit. For all $x \in \mathbb{R}^2 \backslash \{0\}$, the ω-limit set $\omega(x)$ equals this periodic orbit. For $\|x\| < 1$, the α-limit set $\alpha(x) = \{0\}$. ◊

3.5 The Maximal Interval of Existence

As a preparation for the Principal Theorem (Chapter 3.6), we first determine the structure of the maximal domain of definition D for a solution $\Phi : D \to M$.

We consider the vector field $f \in C^1(U, \mathbb{R}^n)$ on the phase space $U \subseteq \mathbb{R}^n$ (or more generally a vector field on a manifold), and we want to assign (for initial time $t_0 = 0$)

[4] For instance, with a norm coming from a Riemannian metric.

to all initial values x_0 from U its maximal interval in time for which the solution to the initial value problem is defined. According to the Picard-Lindelöf theorem, this interval is a neighborhood of 0, and it is open since the same applies to all points on the orbit. So we can write it in the form

$$\big(T^-(x_0), T^+(x_0)\big) \quad \text{with} \quad -\infty \le T^-(x_0) < 0 < T^+(x_0) \le +\infty\,.$$

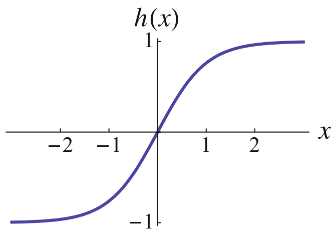
The corresponding solution to the initial value problem is also called the *maximal solution*. We now investigate the *escape times*

$$T^\pm : U \longrightarrow \overline{\mathbb{R}} := \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}\,, \tag{3.5.1}$$

whose values are on the *extended real line*.

To this end, we equip $\overline{\mathbb{R}}$ with a topology that makes $\overline{\mathbb{R}}$ homeomorphic to the interval $[-1, 1]$, the homeomorphism being

$$h : \overline{\mathbb{R}} \to [-1,1],\, x \mapsto \begin{cases} -1 & , x = -\infty \\ \tanh(x) & , x \in \mathbb{R} \\ 1 & , x = +\infty. \end{cases}$$



The following example shows that in general T^+ and T^- are not continuous:

3.36 Example (Escape Times) Consider the initial value problem for the constant vector field $f(x) := e_1$ on the phase space $U := \mathbb{R}^2\backslash\{0\}$; so $\Phi_t(x) = x + e_1 t$. This is defined for all t, provided $x = \binom{x_1}{x_2}$ with $x_2 \neq 0$. But if $x_2 = 0$, then $\big(T^-(x), T^+(x)\big) = \big(-\infty, |x_1|\big)$ if $x_1 < 0$, and $= \big(-x_1, +\infty\big)$ if $x_1 > 0$. ◊

In his example, T^+ jumps upward, but not downward; and this is typical for all differential equations:

3.37 Definition *A function $f : U \to \overline{\mathbb{R}}$ on a topological space*[5] *U is called* **upper semicontinuous** *or, respectively,* **lower semicontinuous** *at $x_0 \in U$, if*

$$f(x_0) \ge \limsup_{x\to x_0} f(x) \quad \textit{or, respectively,} \quad f(x_0) \le \liminf_{x\to x_0} f(x)\,,$$

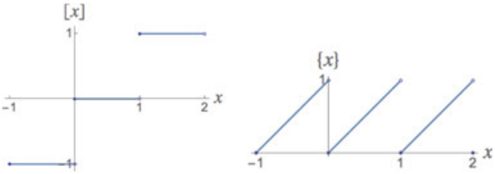
and **upper semicontinuous** *(respectively* **lower semicontinuous***), if it has the respective property at all $x_0 \in U$.*

[5] We will always work in spaces whose topology is metrizable, so we understand

$$\limsup_{x\to x_0} f(x) \quad \text{as} \quad \lim_{\varepsilon \searrow 0} \sup\big\{f(x) \mid x \in U_\varepsilon(x_0) \setminus \{x_0\}\big\}.$$

In general such a function is lower semicontinuous exactly if its epigraph is closed.

3.38 Example (Floor and Ceiling)
When we decompose $x \in \mathbb{R}$ as $x = \lfloor x \rfloor + \{x\}$ with $\lfloor x \rfloor \in \mathbb{Z}$ and $\{x\} \in [0, 1)$, the *floor* function $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{R}$ (also called *Gauss bracket*) is upper semicontinuous, whereas the fractional part $x \mapsto \{x\}$ is lower semicontinuous, and so is the *ceiling* function



$$\lceil \cdot \rceil : \mathbb{R} \to \mathbb{R},$$
$$x \mapsto \min\{z \in \mathbb{Z} \mid z \geq x\}\,. \quad \Diamond$$

3.39 Theorem (Escape Time)
The escape time $T^+ : U \to \overline{\mathbb{R}}$ in (3.5.1) is lower semicontinuous, and the escape time $T^- : U \to \overline{\mathbb{R}}$ is upper semicontinuous. This makes the domain

$$D := \big\{(t, x) \in \mathbb{R} \times U \mid t \in \big(T^-(x), T^+(x)\big)\big\}$$

of $\Phi : D \to U$ (called the **maximal flow***) an open subset of the extended phase space.*

Proof:

- Let $x_0 \in U$. Then there exists, since U is open, a neighborhood $U_r(x_0)$ with $\overline{U_r(x_0)} \subset U$. As $\overline{U_r(x_0)}$ is compact, the restriction of the vector field to this set is Lipschitz continuous. By the Picard-Lindelöf theorem, there exists some $\varepsilon > 0$, such that for all $y \in U_{r/2}(x_0)$, the initial value problem $\dot{x} = f(x),\ x(0) = y$ has a unique solution for $t \in (-\varepsilon, \varepsilon)$.
- Now we consider an increasing sequence of times $(t_n)_{n\in\mathbb{N}}$ with $t_1 = 0, \lim_{n\to\infty} t_n = T^+(x_0)$ such that for appropriate $r_n > 0$ and $\varepsilon_n > 0$, the initial value problems

$$\dot{x} = f(x) \quad , \quad x(0) = y \qquad \text{for all } t \in (-\varepsilon_n, \varepsilon_n) \text{ and } y \in U_{r_n/2}(x_n)$$

  can be solved, where we have set $x_n := \Phi_{t_n}(x_0)$. By construction of the maximal solution we may assume that $t_{n+1} - t_n < \varepsilon_n$.
- Now suppose that T^+ is *not* lower semicontinuous at x_0, i.e., that one has $\hat{T} := \liminf_{x\to x_0} T^+(x) < T^+(x_0)$; then choose $k \in \mathbb{N}$ in such a way that $t_k \leq \hat{T} < t_{k+1}$. By assumption, $t_k + \varepsilon_k > \hat{T}$. By continuous dependence of the flow with respect to the intial conditions (an iteration of Theorem 3.20), there exists a neighborhood $V \subset U$ of x_0 for which $\Phi_{t_k}(V) \subset U_{r_k/2}(x_n)$; and for all $y \in V$, we obtain, in contradiction to the assumption, that

$$T^+(y) = t_k + T^+(\Phi_{t_k}(y)) \geq t_k + \varepsilon_k > \hat{T}\,.$$

- The upper semicontinuity of T^- is shown analogously.

- Now if D were not open, there would be a sequence $(t_n, x_n)_{n\in\mathbb{N}}$ in $(\mathbb{R} \times \mathbb{R}^n) \setminus D$ with limit $(t, x) := \lim_{n\to\infty}(t_n, x_n) \in D$.
  As $U \subseteq \mathbb{R}^n$ is open, we may assume (by dropping the first few terms of the sequence) that $x_n \in U$. Therefore the following alternative applies for each n: Either $t_n \geq T^+(x_n)$ or $t_n \leq T^-(x_n)$. One of the two cases must occur infinitely often, for instance, assume it is $t_n \geq T^+(x_n)$. We pass to the corresponding subsequence. Since $(t, x) \in D$, one has $t \in (0, T^+(x))$; but by

$$\liminf_{n\to\infty} T^+(x_n) \;\leq\; \liminf_{n\to\infty} t_n \;=\; t \;<\; T^+(x)\,,$$

  this is in contradiction to the lower semicontinuity of T^+ at x. □

3.40 Remark (Nonautonomous Differential Equations)
In this section, we have assumed that the differential equation is autonomous. The analogous statements do however hold true for initial value problems of nonautonomous differential equations with a fixed initial time t_0. This is because the ODE can be written in autonomous form by addition of one dependent variable. We will use this fact when we prove our principal theorem. ◊

3.41 Exercise (Escape Time)
The function $H : \mathbb{R} \times (0, \infty) \to \mathbb{R}$, $H(p, q) := \frac{1}{2}p^2 - \frac{m}{q}$ with the parameter $m \in (0, \infty)$ and the corresponding Hamiltonian differential equation $\dot{p} = -m/q^2$, $\dot{q} = p$ describes radial motion in a gravitational field.

(a) Obtain an upper bound for the escape time (i.e., time of collision).
(b) Obtain a lower bound for the escape time.
(c) Find an ODE on $\mathbb{R}$ whose escape times are — depending on the initial value — infinite only in the past, only in the future, or always. ◊

3.6 Principal Theorem of the Theory of Differential Equations

From Theorem 3.20, we know that for continuously differentiable time dependent vector fields f, the solutions $\Phi(t, x_0)$ to the initial value problem

$$\dot{x}(t) = f\big(t, x(t)\big) \quad , \quad x(t_0) = x_0 \tag{3.6.1}$$

depend continuously on the time t *and* on the initial value x_0.

The Principal Theorem will now tell us that the solution Φ is as smooth as f.

The Gronwall inequality, which is used in the proof of the principal theorem, is an important estimate in the theory of differential equations. It is a bit reminiscent of Münchhausen's trick get out of the swamp by pulling himself up at his own hair.

3.42 Theorem **(Gronwall Inequality)**
For $F, G \in C\big([t_0, t_1),\ [0, \infty)\big)$ *and some* $a \geq 0$*, assume the inequality*

$$F(t) \leq a + \int_{t_0}^{t} F(s)G(s)\,\mathrm{d}s \qquad \big(t \in [t_0, t_1)\big). \tag{3.6.2}$$

Then it follows that

$$F(t) \leq a\ \exp\left(\int_{t_0}^{t} G(s)\,\mathrm{d}s\right) \qquad \big(t \in [t_0, t_1)\big). \tag{3.6.3}$$

Proof:
• If $a > 0$, then the right hand side $h(t) := a + \int_{t_0}^{t} F(s)G(s)\,\mathrm{d}s$ satisfies $h(t) > 0$; moreover, due to hypothesis (3.6.2), we have $h'(t) = F(t)G(t) \leq h(t)G(t)$, and therefore $\frac{h'(t)}{h(t)} \leq G(t)$.
By integration one obtains $\ln\left(\frac{h(t)}{a}\right) \leq \int_{t_0}^{t} G(s)\,\mathrm{d}s$, or $h(t) \leq a \exp\left(\int_{t_0}^{t} G(s)\,\mathrm{d}s\right)$. Together with the inequality $F(t) \leq h(t)$, this shows the claim (3.6.3).
• When $a = 0$, the hypothesis as well as the conclusion are satisfied for all $\hat{a} > 0$. Therefore $F = 0$. □

3.43 Remark **(Gronwall Equality)**
The estimate can be memorized easily if one assumes equality. The integral equation

$$F(t) = a + \int_{t_0}^{t} F(s)G(s)\,\mathrm{d}s$$

corresponds to the initial value problem $\dot{F} = F\,G$, $F(t_0) = a$, with the solution $F(t) = a\exp\left(\int_{t_0}^{t} G(s)\,\mathrm{d}s\right)$. ◊

3.6.1 Linearization of the ODE Along a Trajectory

In preparation for the proof of the Principal Theorem, we will first learn which differential equation should be satisfied by the derivative of the solution with respect to the initial value. To this end, we will assume that both the time dependent vector field $f : U \to \mathbb{R}^n$ in (3.6.1) and the solution $\Phi : D \to U$ are continuously differentiable, and for $(t, x) \in D$ we let

$$M(t, x) := \mathrm{D}_2\Phi(t, x) \ \in \mathrm{Mat}(n, \mathbb{R})\,.$$

Then, letting $\tilde{A}(t,x) := \mathrm{D}_2 f(t,x) \in \mathrm{Mat}(n,\mathbb{R})$, one obtains from

$$\Phi(t,x) = x + \int_{t_0}^{t} f\big(s, \Phi(s,x)\big)\,\mathrm{d}s \tag{3.6.4}$$

the integral equation

$$M(t,x) = \mathbb{1} + \int_{t_0}^{t} A(s,x)M(s,x)\,\mathrm{d}s \tag{3.6.5}$$

with $A(s,x) := \tilde{A}\big(s, \Phi(s,x)\big)$. A is continuous, being the composition of continuous maps. We will first study integral equations of type (3.6.5) for arbitrary continuous A. They are equivalent to the linear initial value problem

$$\mathrm{D}_1 M(t,x) = A(t,x)M(t,x) \quad , \quad M(t_0,x) = \mathbb{1}\,. \tag{3.6.6}$$

The continuous dependence of the solution to (3.6.6) on the time t *and* the parameter x, as claimed in the following lemma, does not follow from previously proved statements like Theorem 3.34.

3.44 Lemma
Let $D \subseteq \mathbb{R}_t \times \mathbb{R}^n_x$ be an open neighborhood of (t_0, x_0) that (just as D does in Theorem 3.39) intersects the time axes $\mathbb{R}_t \times \{x\}$ in intervals containing t_0.
Then, for $A \in C\big(D, \mathrm{Mat}(n,\mathbb{R})\big)$, the initial value problem (3.6.6) has a unique solution $M \in C\big(D, \mathrm{Mat}(n,\mathbb{R})\big)$.

Proof:

- By Theorem 3.23, the initial value problem (3.6.6) for initial value $(t_0, \mathbb{1})$ has a unique solution; because x is but a parameter, and A is continuous in t, hence the time dependent vector field is Lipschitz continuous.
- To show continuity of M with respect to (t,x), too, it suffices to restrict the discussion to such compact neighborhoods $K \subset D$ of (t_0, x_0) that are of the form

  $$K := \overline{U_{\delta_t}(t_0)} \times \overline{U_{\delta_x}(x_0)}.$$

  Now $k := \sup_{(t,x)\in K} \|A(t,x)\| < \infty$, hence by (3.6.5)

  $$\|M(t,x)\| \le 1 + k\left|\int_{t_0}^{t} \|M(s,x)\|\,\mathrm{d}s\right| ,$$

  and using Gronwall's lemma (Theorem 3.42),

  $$\sup_{(t,x)\in K} \|M(t,x)\| \le e^{k\delta_t}\,. \tag{3.6.7}$$

- From (3.6.5), we get the identity

$$M(t,x) - M(t,x_0) = \int_{t_0}^{t} A(s,x_0)\big(M(s,x) - M(s,x_0)\big)\,\mathrm{d}s \\ + \int_{t_0}^{t} \big(A(s,x) - A(s,x_0)\big) M(s,x)\,\mathrm{d}s\,. \qquad (3.6.8)$$

  The second term has an upper bound $a(\delta_x) > 0$ that depends on the radius such that $\lim_{\delta_x \searrow 0} a(\delta_x) = 0$. This is because the continuous function A is uniformly continuous on the compact set K, and the norm of M is bounded by (3.6.7).
  Hence we get for $F_x(t) := \|M(t,x) - M(t,x_0)\|$, using (3.6.8):

$$F_x(t) \le a(\delta_x) + \left| \int_{t_0}^{t} k\, F_x(s)\,\mathrm{d}s \right|.$$

  Gronwall's inequality turns this into

$$F_x(t) \le a(\delta_x)\exp(k|t-t_0|) \le a(\delta_x)\exp(k\delta_t) \qquad \big(x \in U_{\delta_x}(x_0)\big),$$

  hence $\lim_{x\to x_0} M(t,x) = M(t,x_0)$ uniformly in $t \in [t_0 - \delta_t, t_0 + \delta_t]$. □

We are now in a position to prove the *Principal Theorem of the theory of ordinary differential equations*.

3.6.2 Statement and Proof of the Principal Theorem

Assume that the time dependent vector field f on extended phase space $U \subseteq \mathbb{R} \times \mathbb{R}^n$ is in $C^r(U, \mathbb{R}^n)$ for some $r \in \mathbb{N}$. We fix an initial time $t_0 \in \mathbb{R}$ and consider, for initial value $(t_0, x_0) \in U$, the maximal interval of existence $\big(T^-(x_0), T^+(x_0)\big)$ for the initial value problem

$$\dot{x} = f(t,x) \quad , \quad x(t_0) = x_0\,. \qquad (3.6.9)$$

Just as in the time independent case, we obtain a maximal domain

$$D = \big\{(t,x) \in U \mid t \in \big(T^-(x), T^+(x)\big)\big\}$$

(open in U) of the nonautonomous flow $\Phi : D \to \mathbb{R}^n$, with

$$\Phi(t_0, x_0) = x_0 \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x_0) = f\big(t, \Phi(t,x_0)\big)\,. \qquad (3.6.10)$$

The nonautonomous flow is as smooth as the vector field:

3.45 Theorem (Principal Theorem in the Theory of ODEs)
If the time dependent vector field f in (3.6.9) satisfies $f \in C^r(U, \mathbb{R}^n)$ for $r \in \mathbb{N}$, then

$$\Phi \in C^r(D, \mathbb{R}^n)\,.$$

Proof:

- From Theorem 3.20, we know that $\Phi \in C^0(D, \mathbb{R}^n)$ when $r = 1$. The time derivative of the solution exists as well, and $D_1\Phi \in C^0(D, \mathbb{R}^n)$; this follows from the second formula in (3.6.10).
  Our first goal is to show that $f \in C^1(U, \mathbb{R}^n)$ implies $\Phi \in C^1(D, \mathbb{R}^n)$ as well. As the time derivative $D_1\Phi$ is continuous, all we need to show is the existence and continuity of the derivative $D_2\Phi : D \to \mathrm{Mat}(n, \mathbb{R})$. If $D_2\Phi$ exists, this mapping has to be continuous, because then $D_2\Phi = M$ where M is the solution to the integral equation

$$M(t, x) = \mathbb{1} + \int_{t_0}^{t} A(s, x)M(s, x)\,\mathrm{d}s\,,$$

  see Lemma 3.44. Continuity of $D_1\Phi$ and $D_2\Phi$ then will imply existence and continuity of $D\Phi$.
  So by definition of the total derivative with respect to x, we have to show, for initial value $(t_0, x_0) \in U$ and times $t \in \bigl(T^-(x_0), T^+(x_0)\bigr)$, that

$$\Phi(t, x_0 + h) - \Phi(t, x_0) = M(t, x_0)h + \mathrm{o}(\|h\|)\,. \tag{3.6.11}$$

- We can Taylor expand the time dependent vector field near (t, x) and obtain

$$f(t, y) = f(t, x) + D_2 f(t, x)(y - x) + R(t, x, y) \tag{3.6.12}$$

  with a remainder $R(t, x, y) = \mathrm{o}(\|y - x\|)$ that is continuous in t, where $t \in [t_0 - \delta_t, t_0 + \delta_t]$.
  Due to $A(s, x_0) = D_2 f\bigl(s, \Phi(s, x_0)\bigr)$ and (3.6.12), we obtain from (3.6.4) and (3.6.5) the following deviation from the linear approximation:

$$\begin{aligned}
&\bigl(\Phi(t, x_0 + h) - \Phi(t, x_0)\bigr) - M(t, x_0)h \\
&= \int_{t_0}^{t} \Bigl(f\bigl(s, \Phi(s, x_0 + h)\bigr) - f\bigl(s, \Phi(s, x_0)\bigr) - A(s, x_0)M(s, x_0)h\Bigr)\,\mathrm{d}s \\
&= \int_{t_0}^{t} D_2 f\bigl(s, \Phi(s, x_0)\bigr) \cdot \Bigl[\Phi(s, x_0 + h) - \Phi(s, x_0) - M(s, x_0)h\Bigr]\,\mathrm{d}s \\
&\quad + \int_{t_0}^{t} R\bigl(s, \Phi(s, x_0), \Phi(s, x_0 + h)\bigr)\,\mathrm{d}s\,.
\end{aligned} \tag{3.6.13}$$

The Gronwall estimate first improves the continuity of Φ to Lipschitz continuity

$$\|\Phi(s, x_0 + h) - \Phi(s, x_0)\| = \mathcal{O}(\|h\|) \qquad \big(s \in [t_0 - \delta_t, t_0 + \delta_t]\big),$$

uniformly on the time interval. Due to the estimate for the remainder, the second term in (3.6.13) has the order $\mathrm{o}\,(\|h\|)$.
Abbreviating

$$F(t) := \|\Phi(t, x_0 + h) - \Phi(t, x_0) - M(t, x_0)h\|,$$

we can write the absolute value of (3.6.13) as

$$F(t) \le \mathrm{o}\,(\|h\|) + k \left| \int_{t_0}^{t} F(s)\,\mathrm{d}s \right| ,$$

with $k := \sup_s \|\mathrm{D}_2 f(s, \Phi(s, x_0))\|$. By Gronwall, we get

$$F(t) = \mathrm{o}\,(\|h\|) e^{k|t-t_0|} = \mathrm{o}\,(\|h\|) \qquad \big(t \in [t_0 - \delta_t, t_0 + \delta_x]\big).$$

This proves (3.6.11).

- To show for $f \in C^r(U, \mathbb{R}^n)$ with $r \ge 2$ that the flow is also r times continuously differentiable, we argue by induction. To this end, we set

$$\tilde{f} : U \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n \quad , \quad \tilde{f}(t, x, h) := \big(f(t, x), \mathrm{D}_2 f(t, x)h\big) .$$

This makes $\tilde{f} \in C^{r-1}(\tilde{U}, \mathbb{R}^n \times \mathbb{R}^n)$ a time dependent vector field on the extended phase space $\tilde{U} := U \times \mathbb{R}^n$. By what we have just shown, and letting $\tilde{D} := D \times \mathbb{R}^n$,

$$\tilde{\Phi} : \tilde{D} \to \mathbb{R}^n \times \mathbb{R}^n \quad , \quad \tilde{\Phi}(t, x_0, h_0) := \big(\Phi(t, x_0), \mathrm{D}_2 \Phi(t, x_0)h_0\big)$$

is a continuous mapping that solves the initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}t}(x, h) = \tilde{f}(t, x, h) \quad , \quad (x, h)(t_0) = (x_0, h_0).$$

We also know that the time derivative $\mathrm{D}_1 \tilde{\Phi}$ is continuous. To show that $\tilde{\Phi} \in C^1(\tilde{D}, \mathbb{R}^n \times \mathbb{R}^n)$, we only need to show the existence and continuity of $\mathrm{D}_2 \tilde{\Phi}$. This is done just like the existence and continuity of $\mathrm{D}_2 \Phi$ was proved above. Using the induction step r times, we obtain $\mathrm{D}^r \Phi \in C^0$, hence $\Phi \in C^r(D, \mathbb{R}^n)$. □

3.6.3 Consequences of the Principal Theorem

We can linearize an autonomous system of differential equations near an equilibrium, but the connection between the solutions of both differential equations is not always that close; see Chapter 7.

The situation near a non-equilibrium is different:

3.46 Theorem (Straightening Theorem)
Let $U \subseteq \mathbb{R}^n$ be open and let the vector field f satisfy $f \in C^r(U, \mathbb{R}^n)$ for some $r \in \mathbb{N}$. Then, if $\tilde{x} \in U$ is not an equilibrium, there exists a C^r-diffeomorphism

$$G : V \to W$$

from an open neighborhood V of $\tilde{x}$ onto some $W \subseteq \mathbb{R}^n$ such that

$$\mathrm{D}G_x f(x) = e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^n \qquad \textit{for all } x \in V .$$

3.47 Remark Therefore, in appropriate coordinates on V, the vector field f is constant, and the local solution to the IVP $\dot{x} = f(x)$, $x(0) = x_0$ is therefore $x(t) = x_0 + e_1 t$, i.e., an affine function of time. ◊

Proof: Since $f(\tilde{x}) \neq 0$, the set

$$F_\varepsilon := \{x \in U_\varepsilon(\tilde{x}) \mid \langle x - \tilde{x}, f(\tilde{x})\rangle = 0\}$$

is, for small $\varepsilon > 0$, a disc of dimension $n - 1$, and

$$\langle f(x), f(\tilde{x})\rangle > 0 \qquad \big(x \in U_\varepsilon(\tilde{x})\big). \tag{3.6.14}$$

By means of a Euclidean transformation T of $\mathbb{R}^n$, i.e., a composition of a translation and a rotation, we can achieve that $T(\tilde{x}) = 0$ and $T\big(f(\tilde{x})\big) = \lambda e_1$ with $\lambda > 0$, namely

$$T(F_\varepsilon) = \big\{y \in \mathbb{R}^n \mid y_1 = 0, \|y\| < \varepsilon\big\} .$$

To simplify notation, we assume that f and $\tilde{x}$ themselves already have the properties $\tilde{x} = 0$ and $f(x_0) = \lambda e_1$ with $\lambda > 0$. Consider the cylinder $Z_\delta := (-\delta, \delta) \times F_\delta$. For small $\delta > 0$, the map $\Phi\restriction_{Z_\delta} : Z_\delta \to U_\varepsilon(x_0)$ is injective; this is because (3.6.14) yields, as long as $\Phi_t(x) \in U_\varepsilon$,

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t, x)\right)_1 = f_1\big(\Phi(t, x)\big) > 0 \, ;$$

therefore no trajectory in U_ε will intersect F_δ more than once. By the Principal Theorem, $\Phi\restriction_{Z_\delta}$ is a C^r-diffeomorphism onto its image $V := \Phi(Z_\delta)$, which is a neighborhood of $\tilde{x}$. Thus the inverse mapping G is also a C^r-diffeomorphism. □

Differential equations frequently depend on parameters $p \in P$, like for instance the mass or the length of a pendulum. We therefore consider the *parametric initial value problem*

$$\dot{x} = f(t, x, p) \quad , \quad x(t_0) = x_0 \tag{3.6.15}$$

with $U \subseteq \mathbb{R} \times \mathbb{R}^n$ and $P \subseteq \mathbb{R}^d$ open and $f \in C^r(U \times P, \mathbb{R}^n)$.

3.48 Theorem *Let $D \subseteq U \times P$ be the maximal domain of the parametric initial value problem (3.6.15). Then the solution Φ satisfies $\Phi \in C^r(D, \mathbb{R}^n)$.*

Proof: The result is immediate if we transition to the initial value problem

$$(\dot{x}, \dot{p}) = \tilde{f}(t, x, p) \quad , \quad (x, p)(t_0) = (x_0, p_0)$$

with the time dependent vector field

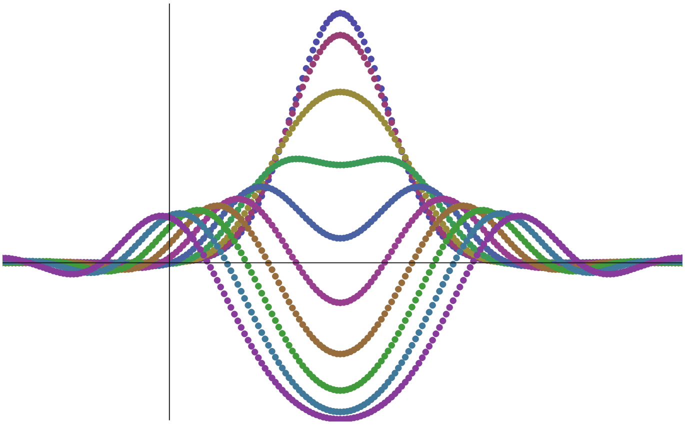
$$\tilde{f} \in C^r\left(\tilde{U}, \mathbb{R}^n \times \mathbb{R}^d\right) \quad , \quad \tilde{f}(t, x, p) := \left(f(t, x, p), 0\right)$$

on $\tilde{U} := U \times P$, because this vector field leaves the value of the parameter invariant and has a solution $\tilde{\Phi} \in C^r(D, \mathbb{R}^{n+d})$. □

The main conclusion from the principal theorem is this one: Near a non-equilibrium, smooth differential equations do *not* possess any *local structure*. All interesting questions are global in nature, i.e., they are questions about the behavior of the solution for large times.

3.49 Literature The books by Amann [AM], Arnol'd [Ar1], Heuser [Heu], Perko [Per] and Walter [Wa1] are standard references for ODE.

Chapter 4
Linear Dynamics



A particularly important class of differential equations consists of the linear ones; by Theorem 3.29, we may assume that we have a system of first order. The *inhomogeneous* initial value problem reads

$$\dot{x}(t) = A(t)\,x(t) + b(t),\, x(t_0) = x_0, \tag{4.0.1}$$

with a *system matrix* $A : I \to \mathrm{Mat}(n, \mathbb{R})$ and a *forcing* $b : I \to \mathbb{R}^n$, t_0 in the interval I, and $x_0 \in \mathbb{R}^n$. The solution strategy consists of first solving the *homogeneous* IVP

$$\dot{y} = A(t)y,\; y(t_0) = x_0, \tag{4.0.2}$$

and then to get the solution of (4.0.1) by means of the *Duhamel principle*, which amounts to an integration. *Formally*, one can write the solution to (4.0.2), in the case of a constant system matrix A, as

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_4

$$y(t) := \exp\big((t - t_0)A\big)x_0 \qquad (t \in \mathbb{R}), \tag{4.0.3}$$

because then, applying the differentiation rule $\exp' = \exp$ to the right hand side, will result in Ay, and clearly $y(t_0) = x_0$.

4.1 Homogeneous Linear Autonomous ODEs

Matrix Exponentials

In order to define and justify Eq. (4.0.3), we will now study matrix valued functions for vector spaces over the scalar fields $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$. Just as in dimension $n = 1$, the exponential function is defined by its power series:

4.1 Definition

- *For an endomorphism* $M \in \mathrm{Lin}(V)$ *of a finite dimensional* $\mathbb{K}$*-vector space* V*, the mapping* $\exp(M) \in \mathrm{Lin}(V)$ *is defined as*

$$\exp(M) := \sum_{k=0}^{\infty} \frac{M^{(k)}}{k!} \tag{4.1.1}$$

  (where $M^{(0)} = \mathrm{Id}_V$ *and* $M^{(k+1)} = M \circ M^{(k)}$*).*
- *The* **matrix exponential** *of matrices in* $\mathrm{Mat}(n, \mathbb{K})$ *is defined in the same way.*

Why does the matrix exponential solve the homogeneous initial value problem (4.0.2) in the case of a constant system matrix? This question is not mere nitpicking, because in the case of a time dependent matrix A, the mapping $t \mapsto \exp\left(\int_{t_0}^{t} A(s)\,\mathrm{d}s\right) x_0$ *does not solve* (4.0.2), except in special cases like dimension $n = 1$.

Equation (4.1.1) involves a series whose terms are elements of $\mathrm{Lin}(V)$, the space of (bounded) linear mappings from V into itself. Suppose for instance that $V = \mathbb{K}^n$, with the Euclidean norm $\|\cdot\|$. The *operator norm* on $\mathrm{Lin}(\mathbb{K}^n)$,

$$\|M\| := \sup_{v \in \mathbb{K}^n \setminus \{0\}} \tfrac{\|M(v)\|}{\|v\|} = \sup_{v \in \mathbb{K}^n,\ \|v\|=1} \|M(v)\|,$$

has the usual properties of a norm, namely $\|M\| \ge 0$, $\|M\| = 0 \Longleftrightarrow M = 0$,

$$\|\lambda M\| = |\lambda|\,\|M\| \quad (\lambda \in \mathbb{K}), \text{ and } \|M + N\| \le \|M\| + \|N\|;$$

moreover, it is *submultiplicative*:

4.2 Lemma *For* $M, N \in \mathrm{Lin}(\mathbb{K}^n)$*, one has the inequality* $\|MN\| \le \|M\|\,\|N\|$.

Proof: Except when $N = 0$ (in which case both sides are 0 anyways), we calculate from $\|MN\| = \sup_{v \neq 0} \frac{\|MNv\|}{\|v\|}$:

$$\|MN\| = \sup_{v:\,Nv\neq 0} \frac{\|MNv\|}{\|Nv\|} \cdot \frac{\|Nv\|}{\|v\|} \leq \sup_{w\neq 0} \frac{\|Mw\|}{\|w\|} \cdot \sup_{v\neq 0} \frac{\|Nv\|}{\|v\|} = \|M\|\,\|N\|\,.$$

□

Based on this lemma, the above definition of $\exp(M)$ is meaningful, because the series converges:

4.3 Lemma (Matrix Exponential) *For* $M \in \mathrm{Lin}(\mathbb{K}^n)$, *the partial sums* $s_k := \sum_{\ell=0}^{k} \frac{M^{(\ell)}}{\ell!}$ $(k \in \mathbb{N})$ *of* $\exp(M)$ *form a Cauchy sequence.*

Proof:
By the triangle inequality and Lemma 4.2, we estimate for $k_1 \geq k_0 \geq \|M\|$:

$$\|s_{k_1} - s_{k_0}\| = \left\| \sum_{\ell=k_0+1}^{k_1} \frac{M^{(\ell)}}{\ell!} \right\| \leq \sum_{\ell=k_0+1}^{k_1} \frac{\|M^{(\ell)}\|}{\ell!} \leq \sum_{\ell=k_0+1}^{k_1} \frac{\|M\|^{\ell}}{\ell!}$$
$$\leq \frac{\|M\|^{k_0+1}}{(k_0+1)!} \sum_{m=0}^{k_1-k_0-1} \frac{\|M\|^m}{(k_0+1)^m} \leq \frac{\|M\|^{k_0+1}}{(k_0+1)!} \left(1 - \frac{\|M\|}{k_0+1}\right)^{-1}.$$

This expression tends to 0 as $k_0 \to \infty$, because the factorial grows faster than the (real variable) exponential function. □

We note in passing that we have not assumed anywhere that the linear endomorphism, or the matrix, is real. This is useful because the Jordan normal form of A in (4.0.2), and thus the one of $\exp(At)$ as well, could be complex.

To see that (4.0.3) indeed solves the initial value problem (4.0.2), we need to be able to differentiate the mapping

$$\mathbb{R} \longrightarrow \mathrm{Lin}(\mathbb{K}^n),\ t \longmapsto \exp(At)$$

with respect to time t. We will also study later how the solutions depend on possible parameters in the linear ODE.

For this kind of question, one evaluates the matrix exponential as a multi-variable function, i.e., one studies the mapping

$$\exp : \mathrm{Lin}(\mathbb{K}^n) \to \mathrm{Lin}(\mathbb{K}^n),\ M \mapsto \exp(M)$$

in its dependence on M. The *Weierstrass test* comes to help:

4.4 Theorem (Weierstrass) *Let* $(V, \|\cdot\|)$ *be a Banach space,* $X \subseteq V$ *and let*

$$f_l : X \to V \qquad (l \in \mathbb{N}_0)$$

be functions with $\sup_{x\in X} \|f_l(x)\| \leq a_l$ *such that* $\sum_{l=0}^{\infty} a_l < \infty$. *Then the series* $\sum_{l=0}^{\infty} f_l$ *converges uniformly on* X.

Proof: The reasoning for the case of series of real valued functions generalizes: Since the metric space V is complete, we have *pointwise* convergence of the partial sums $s_l := \sum_{m=0}^{l} f_m$, i.e.,

$$s(x) := \lim_{l\to\infty} s_l(x) \qquad (x \in X).$$

By hypothesis, there is for every $\varepsilon > 0$ an $m \in \mathbb{N}$ for which $\sum_{l=m+1}^{n} a_l < \varepsilon$ when $n > m$.
This implies $\|s_n(x) - s_m(x)\| \le \sum_{l=m+1}^{n} a_l < \varepsilon \quad (x \in X)$, hence *uniform* convergence on X. □

4.5 Theorem (Exponential Mapping) *For* $A \in \mathrm{Lin}(\mathbb{K}^n)$*, the mapping*

$$\mathbb{R} \to \mathrm{Lin}(\mathbb{K}^n),\, t \mapsto \exp(At)$$

is continuously differentiable, and more specifically,

$$\boxed{\tfrac{\mathrm{d}}{\mathrm{d}t}\exp(At) = A\exp(At).} \tag{4.1.2}$$

Proof: • If we let $V := \mathrm{Lin}(\mathbb{K}^n)$ and $f_l : X \to V,\ M \mapsto \frac{M^{(l)}}{l!}$ in Theorem 4.4, we have for all $l \in \mathbb{N}$: $\sup_{M\in V} \|f_l(M)\| = \infty$.
So we cannot take the entire vector space V as domain X, when using the Weierstrass theorem.
• However, for the ball $X := \{M \in V \mid \|M\| \le r\}$ with radius $r > 0$, one has

$$a_l := \sup_{M\in X} \|f_l(M)\| = \frac{1}{l!}\sup_{M\in X}\left\|M^{(l)}\right\| \le \frac{1}{l!}\sup_{M\in X}\left\|M\right\|^l = \frac{r^l}{l!},$$

and the series $\sum_{l=0}^{\infty} a_l \le \exp(r)$ converges for every choice of r.
• Therefore the exponential mapping is continuous on X, being the uniform limit of continuous functions $s_l\restriction_X : X \to V$; and as the derivatives $\mathrm{D}s_l$ also converge uniformly on X, the exponential mapping is also differentiable. This implies formula (4.1.2), and with it the continuity of the derivative. □

Using the Jordan Normal Form

For the evaluation of $\exp(At)$ in practice, one can use the Jordan normal form of A.

4.6 Definition • *For* $\lambda \in \mathbb{K}$ *and* $r \in \mathbb{N}$*, the matrix* $J_r(\lambda) := \begin{pmatrix} \lambda & 1 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \lambda & 1 \\ 0 & \cdots & 0 & \lambda \end{pmatrix} \in$

$\mathrm{Mat}(r, \mathbb{K})$ *is called an* $r \times r$ **Jordan block with eigenvalue** λ.

- *A **Jordan matrix** is a square matrix of the form*

$$J = \begin{pmatrix} J_{r_1}(\lambda_1) & & & 0 \\ & J_{r_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_{r_k}(\lambda_k) \end{pmatrix} \equiv J_{r_1}(\lambda_1) \oplus \ldots \oplus J_{r_k}(\lambda_k) \,. \qquad (4.1.3)$$

- *A **Jordan basis** of an operator $A \in \mathrm{Lin}(V)$ on the $\mathbb{K}$-vector space V is a basis of V with respect to which the representing matrix A is a Jordan matrix.*

From linear algebra, one has a (constructive) proof of the following theorem:

4.7 Theorem *Let V be a finite dimensional $\mathbb{C}$-vector space and let $A \in \mathrm{Lin}(V)$. Then there exists a Jordan basis for A.*

As the vector space V is isomorphic to $\mathbb{C}^n$ with $n := \dim(V)$, we can write the representing matrix A in the form

$$A = WJW^{-1}, \quad \text{with a Jordan matrix } J \in \mathrm{Mat}(n, \mathbb{C}) \text{ and } W \in \mathrm{GL}(n, \mathbb{C}).$$

Since $\mathrm{Mat}(n, \mathbb{R}) \subset \mathrm{Mat}(n, \mathbb{C})$, we can in particular bring real square matrices A into Jorda n form, but in general, J will not be real; so we are doing a *complexification* of A in the sense of linear algebra.

4.8 Example (Jordan Basis) $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has the complex eigenvalues $\pm\imath$. The matrix $W := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -\imath & \imath \end{pmatrix} \in \mathrm{GL}(2, \mathbb{C})$, with $W^{-1} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & \imath \\ 1 & -\imath \end{pmatrix}$, diagonalizes A:

$$W^{-1}AW = \begin{pmatrix} \imath & 0 \\ 0 & -\imath \end{pmatrix} = J_1(\imath) \oplus J_1(-\imath) \,. \qquad \diamondsuit$$

Just as in this example, it is true in general that for every non-real eigenvalue λ of a real matrix $A \in \mathrm{Mat}\,(n, \mathbb{R})$ and a Jordan block $J_r(\lambda)$, there exists an eigenvalue $\overline{\lambda}$ with the same multiplicity[1] and a Jordan block $J_r(\overline{\lambda})$, because the characteristic polynomial of A can be factored over $\mathbb{R}$ into factors of degree at most two (see Figure 4.1.1).

[1]For $A \in \mathrm{Mat}\,(n, \mathbb{K})$ the *algebraic multiplicity* or simply *multiplicity* of $\lambda \in \mathbb{K}$ is the order of the zero λ of its (monic) characteristic polynomial $p_A \in \mathbb{K}[x]$, $p_A(x) = \det(x\,\mathbb{1}_n - A)$ of degree n, whereas the *geometric multiplicity* is the defect of $A - \lambda\,\mathbb{1}_n$, with $\mathrm{def}(B) = \dim(\ker(B))$.

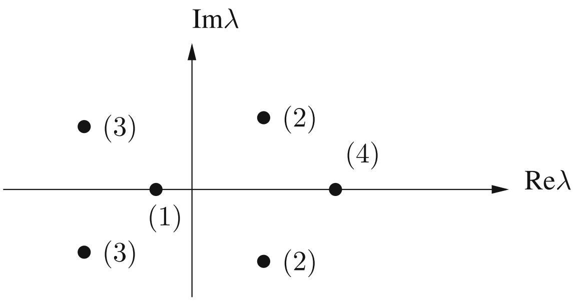


Figure 4.1.1 Complex eigenvalues (with multiplicities) of a real matrix

For $A = WJW^{-1}$, one has $\boxed{\exp(At) = W\exp(Jt)W^{-1} \qquad (t \in \mathbb{R}).}$
This follows from the power series of the matrix exponential by using $A^m = (WJW^{-1})^m = WJ^mW^{-1}$. Since, for a Jordan matrix J like (4.1.3), $\exp(Jt)$ equals

$$\begin{pmatrix} \exp\left(J_{r_1}(\lambda_1)t\right) & & 0 \\ & \ddots & \\ 0 & & \exp\left(J_{r_k}(\lambda_k)t\right) \end{pmatrix} \equiv \exp\bigl(J_{r_1}(\lambda_1)t\bigr) \oplus \ldots \oplus \exp\bigl(J_{r_k}(\lambda_k)t\bigr), \tag{4.1.4}$$

it suffices to calculate $\exp\bigl(J_r(\lambda)t\bigr)$. But $J_r(\lambda) = J_r(0) + \lambda\mathbb{1}$, and $J_r(0)$ and $\lambda\mathbb{1}$ commute; this simplifies the calculation of $\exp\bigl(J_r(\lambda)t\bigr)$:

4.9 Lemma *For commuting matrices* $B, C \in \mathrm{Mat}\,(n, \mathbb{C})$*, i.e., when* $BC = CB$*, one has*

$$\exp(B + C) = \exp(B)\ \exp(C).$$

Proof: Formally, we obtain the identity by plugging $B + C$ into the definition of exp:

$$\begin{aligned} \exp(B+C) &= \sum_{n=0}^{\infty} \frac{1}{n!}(B+C)^n \overset{BC=CB}{=} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} B^i C^{n-i} \\ &= \sum_{n=0}^{\infty} \sum_{i=0}^{n} \frac{1}{i!(n-i)!} B^i C^{n-i} = \left(\sum_{i=0}^{\infty} \frac{1}{i!} B^i\right) \left(\sum_{j=0}^{\infty} \frac{1}{j!} C^j\right), \end{aligned}$$

hence $\exp(B)\ \exp(C)$. This formal calculation is legitimate by a theorem about the Cauchy product, since both series are absolutely convergent. □

4.10 Remarks 1. As a counterexample (in the non-commuting case) we may use the triangular matrices $B := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $C := B^\top = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$: Then $\exp(Bt) =$

$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ and $\exp(Ct) = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$, hence

$$\exp(Bt)\exp(Ct) = \begin{pmatrix} 1+t^2 & t \\ t & 1 \end{pmatrix} \text{ and } \exp(Ct)\exp(Bt) = \begin{pmatrix} 1 & t \\ t & 1+t^2 \end{pmatrix}.$$

In contrast, we infer from $(B+C)^n = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & n \text{ even} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, & n \text{ odd} \end{cases}$ that

$$\exp\big((B+C)t\big) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \sum_{m=0}^{\infty} \frac{t^{2m}}{(2m)!} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \sum_{m=0}^{\infty} \frac{t^{2m+1}}{(2m+1)!} = \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix}. \tag{4.1.5}$$

2. For all $A \in \text{Mat}(n, \mathbb{K})$, we have the *functional equality*

$$\boxed{\exp(At_1)\exp(At_2) = \exp\big(A\,(t_1+t_2)\big) \qquad (t_1, t_2 \in \mathbb{R}),}$$

since multiples of the same matrix commute. The flow Φ_t solving the differential equation $\dot{x} = Ax$ for $A \in \text{Mat}(n, \mathbb{R})$, namely

$$\Phi_t : \mathbb{R}^n \to \mathbb{R}^n,\ \Phi_t(x) = \exp(At)x \qquad (t \in \mathbb{R}),$$

is therefore a one-parameter group. ◊

For the Jordan blocks in (4.1.4), one has $\exp\big(J_r(\lambda)t\big) = \exp\big(J_r(0)t\big)\ \exp\big(\lambda t \mathbb{1}\big)$. Here $\exp(\lambda t \mathbb{1})$ equals $\exp(\lambda t)\mathbb{1}$, and $\exp\big(J_r(0)t\big) = \sum_{n=0}^{\infty} \frac{1}{n!}\big(J_r(0)\big)^n t^n$ with

$$\big(J_r(0)\big)^n_{i,k} = \delta_{i,k-n} \qquad \big(i, k \in \{1, \ldots, r\}\big),$$

hence

$$\exp\big(J_r(\lambda)t\big) = \exp(\lambda t) \begin{pmatrix} 1 & t & \cdots\cdots & \frac{t^{r-1}}{(r-1)!} \\ 0 & 1 & t & \vdots \\ \vdots & \ddots & \ddots\ \ddots & \vdots \\ \vdots & & \ddots\ \ddots & t \\ 0 & \cdots\cdots & 0 & 1 \end{pmatrix}. \tag{4.1.6}$$

4.11 Exercise (Matrix Exponential)

Solve the system of differential equations $\dot{x} = Ax$ with $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$. ◊

Using the Real Jordan Normal Form

For $A \in \mathrm{Mat}\,(n, \mathbb{R})$, it can often be useful to employ the *real* Jordan normal form of $\exp(At)$. It is obtained by setting

$$J_r^{\mathbb{R}}(\lambda) := J_r(\lambda) \qquad (\lambda \in \mathbb{R})$$

for real eigenvalues, and, for $\lambda \in \mathbb{C}\backslash\mathbb{R}$, transforming pairs of Jordan blocks as

$$\begin{pmatrix} J_r(\lambda) & 0 \\ 0 & J_r(\bar{\lambda}) \end{pmatrix},$$

by means of $X := \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbb{1}_r & \mathbb{1}_r \\ -\imath\, \mathbb{1}_r & \imath\, \mathbb{1}_r \end{pmatrix}$ into the *real normal form*

$$J_r^{\mathbb{R}}(\lambda) := X \begin{pmatrix} J_r(\lambda) & 0 \\ 0 & J_r(\bar{\lambda}) \end{pmatrix} X^{-1} = \begin{pmatrix} J_r(\mu) & -\varphi\, \mathbb{1}_r \\ \varphi\, \mathbb{1}_r & J_r(\mu) \end{pmatrix} \qquad (\lambda \in \mathbb{C}\backslash\mathbb{R}),$$

where $\mu := \mathrm{Re}(\lambda)$ and $\varphi := \mathrm{Im}(\lambda)$. Now $\exp\left(J_r^{\mathbb{R}}(\lambda)t\right)$ equals

$$X \exp\left(\begin{pmatrix} J_r(\lambda) & 0 \\ 0 & J_r(\bar{\lambda}) \end{pmatrix} t\right) X^{-1} = e^{\mu t} \begin{pmatrix} \cos(\varphi t) e^{J_r(0)t} & -\sin(\varphi t) e^{J_r(0)t} \\ \sin(\varphi t) e^{J_r(0)t} & \cos(\varphi t) e^{J_r(0)t} \end{pmatrix}, \tag{4.1.7}$$

and in the special case of multiplicity $r = 1$, one has

$$\exp\left(J_1^{\mathbb{R}}(\lambda)t\right) = e^{\mu t} \begin{pmatrix} \cos(\varphi t) & -\sin(\varphi t) \\ \sin(\varphi t) & \cos(\varphi t) \end{pmatrix}. \tag{4.1.8}$$

Interpretation of the Trace

The image $g(\Lambda)$ of a compact set $\Lambda \subset \mathbb{R}^n$ under a diffeomorphism $g : \mathbb{R}^n \to \mathbb{R}^n$ has the volume

$$V_n\big(g(\Lambda)\big) = \int_\Lambda \big|\det\big(\mathrm{D}g(x)\big)\big|\, \mathrm{d}x\,,$$

so the absolute value of the functional determinant at $x \in \Lambda$ is the factor by which g enlarges the volume near x; this is the *transformation theorem*.

Let us now consider the solution operator for the differential equation $\dot{x} = Ax$ for time $t \in \mathbb{R}$, namely the linear mapping

$$\Phi_t \in \mathrm{Lin}(\mathbb{R}^n),\ \Phi_t(x) = \exp(At)x\,.$$

The derivative is constant, as with every linear mapping:

$$\mathrm{D}\Phi_t(x) = \exp(At) \qquad (x \in \mathbb{R}^n), \tag{4.1.9}$$

and we have the

4.12 Theorem *For* $A \in \mathrm{Lin}(\mathbb{C}^n)$*, one has* $\boxed{\det\big(\exp(A)\big) = \exp\big(\mathrm{tr}(A)\big).}$

Proof: This is a consequence of the existence of a Jordan basis, i.e., of the existence of $W \in \mathrm{GL}(n, \mathbb{C})$ such that $W^{-1}AW = J$ with a Jordan matrix J. Using the notation (4.1.3) for J, one has

$$\begin{aligned}\det\big(\exp(A)\big) &= \det\big(W^{-1}\exp(A)W\big) = \det\big(\exp(W^{-1}AW)\big) = \det\big(\exp(J)\big) \\ &= \textstyle\prod_{\ell=1}^{k} \det\big(\exp(J_{r_\ell}(\lambda_\ell))\big) = \prod_{\ell=1}^{k} \exp(r_\ell \lambda_\ell)\end{aligned}$$

and on the other hand $\mathrm{tr}(A) = \mathrm{tr}(W^{-1}AW) = \mathrm{tr}(J) = \sum_\ell \mathrm{tr}\big(J_{r_\ell}(\lambda_\ell)\big) = \sum_\ell r_\ell \lambda_\ell$, so the claim follows from the functional equation of the complex exponential function. □

Conclusion: The linear flow $\Phi_t(x) = \exp(At)x$ generated by $A \in \mathrm{Mat}(n, \mathbb{R})$ on the phase space $\mathbb{R}^n$ is volume preserving if and only if $\mathrm{tr}(A) = 0$.

4.13 Remark (**Real General Linear Group**)
$\mathrm{GL}(n, \mathbb{R})$, the group of invertible matrices in $\mathrm{Mat}(n, \mathbb{R})$, is the most important example of a Lie group. $\mathrm{GL}(n, \mathbb{R})$ is the open and dense subset consisting of those matrices A in the n^2-dimensional vector space $\mathrm{Mat}(n, \mathbb{R})$ for which $\det(A) \neq 0$. This makes $\mathrm{GL}(n, \mathbb{R})$ into a submanifold of $\mathrm{Mat}(n, \mathbb{R})$, see Definition 2.34. Matrix multiplication and inversion are smooth operations on the entries of the matrix (in the latter case, this follows from Cramer's rule in Linear Algebra).
If we consider the mapping

$$\mathrm{Mat}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R}),\ u \mapsto \exp(u),$$

the image of the exponential function is indeed contained in $\mathrm{GL}(n, \mathbb{R})$, because

$$\det\big(\exp(u)\big) = \exp\big(\mathrm{tr}(u)\big) > 0\,.$$

As there do exist $g \in \mathrm{GL}(n, \mathbb{R})$ with $\det(g) < 0$, it is clear that exp is not surjective; rather the image is the subgroup $\mathrm{GL}^+(n, \mathbb{R})$, see Example E.18).

On the other hand, the exponential map is invertible at least for those $A \in \mathrm{GL}(n, \mathbb{R})$ that satisfy $\|\mathbb{1} - A\| < 1$, because this is where the power series of the inverse function converges:

$$\ln(A) = \ln\big(\mathbb{1} - (\mathbb{1} - A)\big) = -\sum_{k=1}^{\infty} \frac{(\mathbb{1} - A)^k}{k}\,.$$

$\mathrm{Mat}(n, \mathbb{R})$ together with the commutator is an example of a Lie Algebra.[2] The commutator, $[u_1, u_2] = u_1 u_2 - u_2 u_1$ if $u_1, u_2 \in \mathrm{Mat}(n, \mathbb{R})$, measures the lack of commutativity in the group multiplication, because

$$\exp(\varepsilon u_1)\exp(\varepsilon u_2)\exp(-\varepsilon u_1)\exp(-\varepsilon u_2) = \mathbb{1} + \varepsilon^2[u_1, u_2] + \mathcal{O}(\varepsilon^3)\,.$$

This is why $\mathrm{Mat}(n, \mathbb{R})$ is called *the Lie algebra of* $\mathrm{GL}(n, \mathbb{R})$. ◊

4.2 Explicitly Time Dependent Linear ODEs

The Homogeneous Problem

First we consider, on a time interval I containing t_0, the linear *homogeneous*, but non-autonomous (i.e., explicitly time dependent) initial value problem

$$\dot{y}(t) = A(t)y(t),\ y(t_0) = y_0, \tag{4.2.1}$$

where $A : I \to \mathrm{Mat}(n, \mathbb{R})$ is a continuous matrix valued function.

4.14 Theorem *The initial value problem (4.2.1) has a unique solution*

$$y : I \to \mathbb{R}^n \text{ with } \|y(t)\| \le \exp\left(\left|\int_{t_0}^t \|A(s)\|\,\mathrm{d}s\right|\right)\|y_0\|\,. \tag{4.2.2}$$

Proof:

- It suffices to consider $t \ge t_0$ because (4.2.2) is invariant under time reversal.
- The unique solvability of the initial value problem for the time dependent vector field $(t, x) \mapsto A(t)x$ on the interval I follows from Theorem 3.23 with the time dependent Lipschitz constant $L(t) := \|A(t)\|$.
- For $t \in I$, the function $z(t) := \exp\left(-\int_{t_0}^t \|A(s)\|\,\mathrm{d}s\right) y(t)$ satisfies the initial value problem

$$\dot{z}(t) = N(t)z(t),\ z(0) = y_0 \text{ with } N(t) := A(t) - \|A(t)\|\,\mathbb{1},$$

  as can be checked by differentiating. Estimate (4.2.2) follows once we have proved

$$\|z(t)\| \le \|y_0\| \qquad (t \ge t_0, t \in I). \tag{4.2.3}$$

[2]**Definition:** A *Lie Algebra* is a vector space E together with a bilinear anti-commutative mapping $[\cdot,\cdot] : E \times E \to E$, that satisfies the *Jacobi identity* (see also Appendix E.3)

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0\,.$$

Now for $S(t) := N(t) + N^\top(t)$ and $t \geq t_0$, we calculate

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\|z(t)\|^2 &= \langle \dot z(t), z(t)\rangle + \langle z(t), \dot z(t)\rangle = \langle N(t)z(t), z(t)\rangle + \langle z(t), N(t)z(t)\rangle \\ &= \langle z(t), S(t)z(t)\rangle \leq 0,\end{aligned} \qquad (4.2.4)$$

because $S(t)$ is self-adjoint and has only non-positive eigenvalues E.
The latter can be seen as follows: If it were true that $S(t)v = Ev$ with $v \in \mathbb{R}^n \setminus \{0\}$ and $E > 0$, then it would also follow $(A(t) + A(t)^\top)v = (E + 2\|A(t)\|)v$, hence

$$\|A(t) + A(t)^\top\| > 2\|A(t)\| = \|A(t)\| + \|A(t)^\top\|,$$

in contradiction to the triangle inequality for the operator norm. But (4.2.4) implies (4.2.3).

□

If $\varphi_1, \varphi_2 : I \to \mathbb{R}^n$ are solutions to the ODE $\dot y(t) = A(t)y(t)$ and $c_1, c_2 \in \mathbb{R}$, then the linear combination $c_1\varphi_1 + c_2\varphi_2 : I \to \mathbb{R}^n$ also solves the differential equation. The set

$$L_0 := \left\{\varphi \in C^1(I, \mathbb{R}^n) \mid \dot\varphi(t) = A(t)\varphi(t),\ t \in I\right\} \qquad (4.2.5)$$

of solutions forms therefore an $\mathbb{R}$-subspace of $C^1(I, \mathbb{R}^n)$, called the (*homogeneous*) *solution space*. If $t_0 \in I$, then by local existence and uniqueness of the solution to the initial value problem, the linear mapping

$$B_{t_0} : L_0 \to \mathbb{R}^n, \varphi \mapsto \varphi(t_0)$$

is an isomorphism, hence $\dim(L_0) = n$.

4.15 Definition *A basis of the homogeneous solution space L_0 is called a* ***fundamental system*** *of solutions to the differential equation.*

As the solution to the homogeneous system (4.2.1) depends linearly on y_0, we can write the general solution in the form

$$y_h(t) = \Phi(t, s)y_h(s),$$

where $\Phi : I \times I \to \mathrm{Mat}(n, \mathbb{R})$ is called the *solution operator*. In the case of a time independent matrix A,

$$\Phi(t, s) = \exp\big((t - s)\,A\big) \qquad (t, s \in \mathbb{R}).$$

In the general case, one has

$$\boxed{\tfrac{\mathrm{d}}{\mathrm{d}t}\Phi(t, s) = A(t)\Phi(t, s) \text{ and } \Phi(s, s) = \mathbb{1}.} \qquad (4.2.6)$$

4.16 Remarks 1. Even though this family of matrices depends on *two* parameters, namely the initial time s and the final time t, it suffices to know the one-parameter family

$$t \mapsto \Phi(t, t_0) \qquad (t \in I)$$

for a single $t_0 \in I$ because $\Phi(t,s) = \Phi(t,t_0)\,\Phi(t_0,s) = \Phi(t,t_0)\,\Phi(s,t_0)^{-1}$.

2. The solution of such a homogeneous nonautonomous problem will often be a matrix whose entries are not elementary functions, even if the entries of A are elementary functions.
As a matter of fact, many so-called *higher transcendental functions* are defined as solutions to this kind of ODEs, for instance the Bessel equation and the Mathieu equation. ♢

The Wronskian

4.17 Definition *For an interval I and curves $v_1, \ldots, v_n \in C(I, \mathbb{R}^n)$, their* ***Wronskian****, or* ***Wronski determinant****, is the continuous function*

$$w : I \to \mathbb{R},\, t \mapsto \det\big(v_1(t), \ldots, v_n(t)\big)\,.$$

We are interested in Wronskians w of solutions $v_1, \ldots, v_n$ to the ODE (4.2.1). As these solutions are C^1 functions of time, the Wronskian satisfies $w \in C^1(I, \mathbb{R})$ and also solves a certain differential equation:

4.18 Theorem *The Wronskian w satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}t} w(t) = \mathrm{tr}\big(A(t)\big)\, w(t) \qquad (t \in I), \tag{4.2.7}$$

and therefore

$$w(t) = \exp\left(\int_{t_0}^{t} \mathrm{tr}\big(A(s)\big)\,\mathrm{d}s\right) w(t_0)\,. \tag{4.2.8}$$

Proof:
Since the solutions $v_i \in C^1(I, \mathbb{R}^n)$ to the differential equation have the derivative

$$\dot{v}_i(t) = A(t)v_i(t) \qquad (i = 1, \ldots, n),$$

it follows from the product rule that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} w(t) &= \sum_{i=1}^{n} \det\big(v_1(t), \ldots, v_{i-1}(t), \dot{v}_i(t), v_{i+1}(t), \ldots, v_n(t)\big)\\ &= \sum_{i=1}^{n} \det\big(v_1(t), \ldots, v_{i-1}(t), Av_i(t), v_{i+1}(t), \ldots, v_n(t)\big)\,.\end{aligned}$$

For the canonical basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$, one has

$$\operatorname{tr}(A) = \sum_{i=1}^{n} (A)_{i,i} = \sum_{i=1}^{n} \det\bigl(e_1, \ldots, e_{i-1}, Ae_i, e_{i+1}, \ldots, e_n\bigr),$$

and this generalizes to arbitrary vectors $\tilde{v}_1, \ldots, \tilde{v}_n \in \mathbb{R}^n$ in view of the product rule for determinants:

$$\begin{aligned}
\operatorname{tr}(A) \det(\tilde{v}_1, \ldots, \tilde{v}_n) &= \sum_{i=1}^{n} \det\bigl((\tilde{v}_1, \ldots, \tilde{v}_n)(e_1, \ldots, e_{i-1}, Ae_i, e_{i+1}, \ldots, e_n)\bigr) \\
&= \sum_{i=1}^{n} \sum_{k=1}^{n} \det\bigl(\tilde{v}_1, \ldots, \tilde{v}_{i-1}, (A)_{k,i}\tilde{v}_k, \tilde{v}_{i+1}, \ldots, \tilde{v}_n\bigr) \\
&= \sum_{i=1}^{n} \det\bigl(\tilde{v}_1, \ldots, \tilde{v}_{i-1}, A\tilde{v}_i, \tilde{v}_{i+1}, \ldots, \tilde{v}_n\bigr),
\end{aligned}$$

hence the ODE (4.2.7). Its solution (4.2.8) is obtained by separation of variables: $\log\left(\frac{w(t)}{w(t_0)}\right) = \int_{w(t_0)}^{w(t)} \frac{\mathrm{d}u}{u} = \int_{t_0}^{t} \operatorname{tr}\bigl(A(s)\bigr)\,\mathrm{d}s$ if $w(t_0) \in \mathbb{R}\backslash\{0\}$. □

4.19 Remarks (Wronskian)

1. So the Wronskian of a *fundamental system* can be calculated by integration, even if only the initial values $v_1(t_0),\ldots, v_n(t_0)$ are known. And since $w(t_0) = \det\bigl(v_1(t_0), \ldots, v_n(t_0)\bigr) \neq 0$, it also follows from Formula (4.2.8) for the solution that $w(t) \neq 0$ for all $t \in I$.
2. The ratio $w(t)/w(t_0)$ of the Wronskian of a fundamental system gives the factor by which the volume of $\Phi(t, t_0)(K)$ at time t has changed compared to the volume of the compact $K \subset \mathbb{R}^n$.
3. For a linear differential equation of n^{th} order in the form

$$y^{(n)}(t) + \sum_{i=0}^{n-1} a_i(t) y^{(i)}(t) = 0$$

with continuous coefficients $a_0, \ldots, a_{n-1}$, the corresponding differential equation of first order according to Theorem 3.29 has the form

$$\begin{pmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ with } x_i = y^{(i-1)} \text{ and } A = \begin{pmatrix} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ -a_0 & -a_1 & \ldots & -a_{n-1} \end{pmatrix}.$$

In this case the Wronskian of solutions $y_1, \dots, y_n : I \to \mathbb{R}$ equals

$$w(t) = \det \begin{pmatrix} y_1(t) & \dots & y_n(t) \\ \vdots & & \vdots \\ y_1^{(n-1)}(t) & \dots & y_n^{(n-1)}(t) \end{pmatrix}, \text{ and } \tfrac{d}{dt} w(t) = -a_{n-1}(t) w(t)\,. \qquad \diamondsuit$$

The Inhomogeneous Problem

The *inhomogeneous initial value problem* with continuous forcing $b : I \to \mathbb{R}^n$,

$$\dot{z}(t) = A(t)z(t) + b(t),\, z(t_0) = z_0 \tag{4.2.9}$$

can easily be solved if the homogeneous solution operator from (4.2.6) is known.

4.20 Theorem (**'Duhamel Principle'**)
The solution of the initial value problem (4.2.9) is

$$\boxed{z(t) = \Phi(t, t_0)z_0 + \textstyle\int_{t_0}^{t} \Phi(t, s)b(s)\, ds \qquad (t \in I).} \tag{4.2.10}$$

Proof: • From $\Phi(s, s) = 1\!\!1$, one gets $z(t_0) = z_0$ in (4.2.10).
• Moreover, from $\frac{d}{dt}\Phi(t, s) = A(t)\Phi(t, s)$, one gets

$$\dot{z}(t) = A(t)\Phi(t, t_0)z_0 + A(t)\int_{t_0}^{t} \Phi(t, s)b(s)\, ds + \Phi(t, t)b(t) = A(t)z(t) + b(t).$$

Therefore, (4.2.10) is the solution to (4.2.9). □

The set of solutions to the inhomogeneous linear differential equation,

$$L_b := \{\varphi \in C^1(I, \mathbb{R}^n) \mid \dot{\varphi}(t) = A(t)\varphi(t) + b(t), t \in I\},$$

has the form

$$L_b = L_0 + \varphi_b,$$

where L_0 is the homogeneous solution space defined in (4.2.5) and φ_b is the particular solution

$$\varphi_b \in C^1(I, \mathbb{R}^n),\, \varphi_b(t) = \int_{t_0}^{t} \Phi(t, s)b(s)\, ds.$$

This makes the *inhomogeneous solution space* L_b into an n-dimensional affine subspace of $C^1(I, \mathbb{R}^n)$.

4.21 Example (**Inhomogeneous Problem**)

The ODE $\ddot{x}(t) + \frac{1}{10}\dot{x}(t) + x(t) = \cos(t)$ of a damped harmonic oscillator with external forcing is equivalent to the system

$$\dot{z}(t) = A(t)z(t) + b(t)$$

with $z(t) := \begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix}$ and $A(t) = \begin{pmatrix} 0 & 1 \\ -1 & -1/10 \end{pmatrix}$, $b(t) = \begin{pmatrix} 0 \\ \cos(t) \end{pmatrix}$. A has the complex eigenvalues

$$\lambda_{1/2} = -\tfrac{1}{20} \pm \sqrt{\left(\tfrac{1}{20}\right)^2 - 1} = \tfrac{1}{20}(-1 \pm \sqrt{399}\,\imath)$$

and eigenvectors

$$W_{1/2} = \begin{pmatrix} \frac{1}{20}(+1 \mp \sqrt{399}\imath) \\ 1 \end{pmatrix};$$

so it follows, with the diagonalizing matrix

$$W := (W_1; W_2) = \begin{pmatrix} \frac{1-\sqrt{399}\imath}{20} & \frac{1+\sqrt{399}\,\imath}{20} \\ 1 & 1 \end{pmatrix}, \text{ hence } W^{-1} = \begin{pmatrix} \frac{-10\imath}{\sqrt{399}} & \frac{1}{2} - \frac{\imath}{2\sqrt{399}} \\ \frac{10\imath}{\sqrt{399}} & \frac{1}{2} + \frac{\imath}{2\sqrt{399}} \end{pmatrix},$$

that $W^{-1}AW = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Consequently, with $\omega := \frac{\sqrt{399}}{20}$, one gets

$$\exp(At) = W \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} W^{-1} = e^{-t/20} \begin{pmatrix} \cos(\omega t) + \frac{\sin(\omega t)}{\sqrt{399}} & \frac{1}{\omega}\sin(\omega t) \\ -\frac{1}{\omega}\sin(\omega t) & \cos(\omega t) - \frac{\sin\left(\frac{\sqrt{399}}{20}t\right)}{\sqrt{399}} \end{pmatrix}.$$

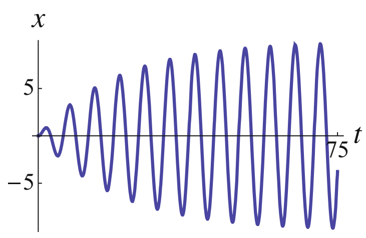
From (4.2.10), one obtains

$$z(t) = \exp(At)z_0 + \int_0^t \exp\big(A(t-s)\big)b(s)\,ds =$$
$$e^{-t/20} \begin{pmatrix} \cos(\omega t) + \frac{\sin(\omega t)}{\sqrt{399}} & \frac{1}{\omega}\sin(\omega t) \\ -\frac{1}{\omega}\sin(\omega t) & \cos(\omega t) - \frac{\sin(\omega t)}{\sqrt{399}} \end{pmatrix} z_0 + 10 \begin{pmatrix} \sin(t) - \frac{1}{\omega}e^{-t/20}\sin(\omega t) \\ \cos(t) + e^{-t/20}\left(-\cos(\omega t) + \frac{\sin(\omega t)}{\sqrt{399}}\right) \end{pmatrix}.$$

Therefore, the general solution to this second order ODE has the form

$$x(t) = e^{-t/20}\big(c_1\cos(\omega t) - c_2\sin(\omega t)\big) + 10\sin t.$$

The solution to the initial value problem for $x_0 = x_0' = 0$ is depicted on the right.



While there are no difficulties with the integration in principle, it is already calculationally complex in this simple example. ◊

4.3 Quasipolynomials

The method of quasipolynomials is used to reduce the calculational effort when solving a linear inhomogeneous differential equation with constant coefficients.
Reminder: The solution operator $\exp(At)$ for the linear ODE $\dot{x} = Ax$ is

$$\exp(At) = W \exp(Jt) W^{-1}$$

with the Jordan matrix $J = \exp\big(J_{r_1}(\lambda_1)\big) \oplus \ldots \oplus \exp\big(J_{r_k}(\lambda_k)\big)$, and according to (4.1.6), one has $\exp\big(J_r(\lambda)t\big) = e^{\lambda t} \begin{pmatrix} 1 & t & \cdots & \cdots & \frac{t^{r-1}}{(r-1)!} \\ 0 & 1 & t & & \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & t \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$.

Conclusion: If $A \in \mathrm{Mat}(n, \mathbb{C})$, and $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ are the eigenvalues of A with multiplicities $\nu_1, \ldots, \nu_k$, then the entries of the matrix $\exp(At)$ are of the form $\sum_{l=1}^{k} e^{\lambda_l t}\, p_l(t)$, where $p_l(t)$ is a polynomial of degree $\leq \nu_l - 1$.
We can use this conclusion to find a solution for linear ODEs directly by using a corresponding ansatz for the solution.

4.22 Definition *For* $\lambda \in \mathbb{K}$ *and a polynomial* $p \in \mathbb{K}[t]$, *we call the function* $t \mapsto e^{\lambda t} p(t)$ *a* λ-**quasipolynomial** *of* **degree** $\deg(p)$ over $\mathbb{K}$.

If the eigenvalues $\lambda_1, \ldots, \lambda_k$ and their algebraic multiplicities $\nu_1, \ldots, \nu_k$ are known from evaluation of the characteristic polynomial of A, one can make a solution ansatz in the form given above.

The $\mathbb{K}$-vector space of λ-quasipolynomials is mapped into itself under differentiation, and we have the formula

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\big(e^{\lambda t} p(t)\big) = e^{\lambda t}\big(p'(t) + \lambda p(t)\big).$$

Plugging the solution ansatz into the ODE yields, for each eigenvalue λ_l, an equation for the polynomials p_l (in general a system of equations because $x(t) = \big(x_1(t), \ldots, x_n(t)\big)^\top$). In principle, the coefficients can be calculated from this.

The solution ansatz is particularly simple in the case of single differential equations of higher order. For if

$$x^{(n)} + a_{n-1} x^{(n-1)} + \ldots + a_0 x = 0,$$

then the equivalent system $\dot{y} = Ay$ with $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $y_k := \frac{d^{k-1}x}{dt^{k-1}}$

has the matrix $A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{n-2} & -a_{n-1} \end{pmatrix}$.

The characteristic polynomial of A is then simply

$$\det(\lambda \mathbb{1} - A) = \det \begin{pmatrix} \lambda & -1 & 0 & \dots & 0 \\ 0 & \lambda & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & & \lambda & -1 \\ a_0 & a_1 & \dots & a_{n-2} & \lambda + a_{n-1} \end{pmatrix} = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_0,$$

so its coefficients are just the coefficients of the differential equation.

So we do not need to make the detour to a first order system when we want to write the general solution in the form of a quasipolynomial.

4.23 Examples (**Quasipolynomials for ODE**)

1. $x^{(4)} - ax = 0$, $a > 0$. The zeros of the characteristic polynomial $\lambda^4 - a = 0$ are

$$\lambda_k = \imath^k \sqrt[4]{a} \qquad (k = 1, \dots, 4).$$

Therefore, each solution is of the form $x(t) = \sum_{k=1}^{4} c_k e^{\lambda_k t}$ with coefficients $c_k \in \mathbb{C}$. If we are looking for a real solution, we obviously need $c_2, c_4 \in \mathbb{R}$ and $c_1 = \overline{c}_3$. Thus, with $\omega := \sqrt[4]{a} > 0$ the general real solution is of the form

$$x(t) = d_1 \exp(\omega t) + d_2 \exp(-\omega t) + d_3 \cos(\omega t) + d_4 \sin(\omega t) \qquad (d_k \in \mathbb{R}).$$

2. The differential equation $\ddot{x} + k\dot{x} + x = 0$ with $k > 0$ (see Example 4.21) describes a *damped harmonic oscillator* (without external forcing).

The two eigenvalues $\lambda_{1/2} = -\frac{k}{2} \pm \sqrt{\frac{k^2}{4} - 1}$ of the matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & -k \end{pmatrix}$ only coincide in the *critically damped case* $k = 2$: In this case $\lambda_1 = \lambda_2 = -1$, so the general solution is $x(t) = (c_1 + c_2 t)e^{-t}$ (see also Chapter 5.4). ◊

For $\lambda \in \mathbb{R}$ (even for $\lambda \in \mathbb{C}$!) we have $\cosh(\lambda t) = \frac{1}{2}\left(e^{\lambda t} + e^{-\lambda t}\right)$ and $\sinh(\lambda t) = \frac{1}{2}\left(e^{\lambda t} - e^{-\lambda t}\right)$, and thus by Euler's formula, $\cos(\lambda t) = \frac{1}{2}\left(e^{\imath\lambda t} + e^{-\imath\lambda t}\right)$ and $\sin(\lambda t) = \frac{1}{2\imath}\left(e^{\imath\lambda t} - e^{-\imath\lambda t}\right)$. Solutions to linear ODEs with constant coefficients can therefore be written in terms of products of these four elementary functions and powers of t, because this is what one obtains by linear combinations of appropriate quasipolynomials.

Note a further consequence: If $\dot{x} = Ax + b(t)$, where $b(t)$ is a sum of quasipolynomials, then the solution to this *inhomogeneous* differential equation with constant coefficients can be written as a sum of λ-quasipolynomials (where λ is either an eigenvalue of A or an exponent that occurs in $b(t)$). This is an immediate consequence of the solution formula for the initial value problem with $\varphi(0) = x_0$ in this case (see (4.2.10)):

$$\varphi(t) = \exp(At)\left(x_0 + \int_0^t \exp(-As)b(s)\,\mathrm{d}s\right),$$

because products and integrals of quasipolynomials are quasipolynomials.

4.24 Examples (**Inhomogeneous Linear Differential Equations**)

1. $\ddot{x} + x = t^2$. A particular solution to this inhomogeneous differential equation is $x_p(t) := t^2 - 2$, so the general solutions is $a_1 \cos t + a_2 \sin t + x_p(t)$, with $a_1, a_2 \in \mathbb{R}$.
2. $x^{(4)} + x = t^2 e^t \cos t$. The right hand side is of the form $\frac{1}{2}t^2\left(e^{\lambda t} + e^{\overline{\lambda} t}\right)$ with $\lambda := 1 + \imath$.

   Generally speaking, the k^{th} derivative of a quasipolynomial $e^{\lambda t} p(t)$ is

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left(e^{\lambda t} p(t)\right) = e^{\lambda t} \sum_{l=0}^{k} \binom{k}{l} \lambda^l p^{(k-l)}(t), \tag{4.3.1}$$

   because under the Leibniz rule, there are $\binom{k}{l}$ possibilities to hit the exponential factor with l derivatives.

   So for a particular solution x_p, we make the following ansatz: $x_p = y_p + \overline{y}_p$ with $y_p^{(4)}(t) + y_p(t) = \frac{t^2}{2} e^{\lambda t}$, where $y_p(t) := e^{\lambda t}(a_2 t^2 + a_1 t + a_0)$. By formula (4.3.1), the left hand side $y_p^{(4)}(t) + y_p(t) =$

$$e^{\lambda t}\left((\lambda^4 + 1)a_2 t^2 + [(\lambda^4 + 1)a_1 + 8\lambda^3 a_2]t + [(\lambda^4 + 1)a_0 + 4\lambda^3 a_1 + 12\lambda^2 a_2]\right).$$

   Comparing with the right hand side and using $\lambda^2 = 2\imath,\ \lambda^4 = -4$, we get

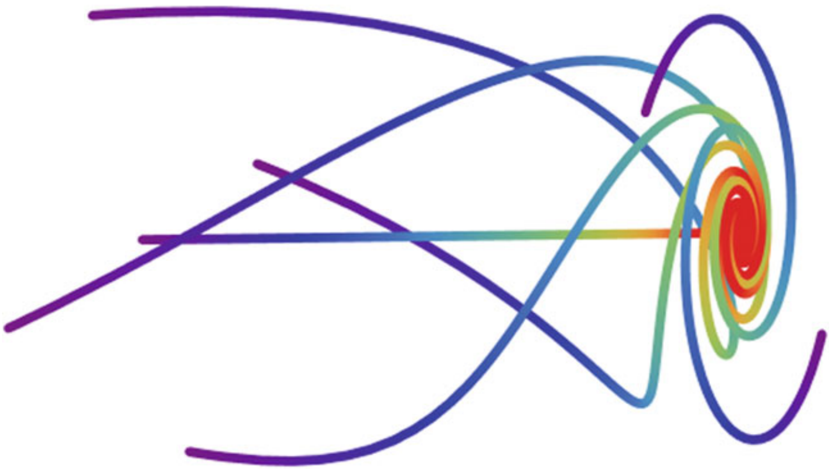
$$a_2 = \frac{1/2}{\lambda^4+1} = -\frac{1}{6}, a_1 = \frac{-8\lambda^3 a_2}{\lambda^4+1} = \frac{8}{9}(1-\imath) \text{ and } a_0 = -\frac{4\lambda^3 a_1 + 12\lambda^2 a_2}{\lambda^4+1} = \frac{92}{27}\imath.$$

   Thus we find $x_p(t) = 2\mathrm{Re}\left(e^{(1+\imath)t}(-\frac{t^2}{6} + \frac{8}{9}(1-\imath)t + \frac{92}{27}\imath)\right) =$

$$e^t\left((-\frac{t^2}{3} + \frac{16}{9}t)\cos t + (\frac{16}{9}t - \frac{184}{27})\sin t\right).$$

◇

Chapter 5
Classification of Linear Flows



We know the flow on the phase space $\mathbb{R}^n$ that is generated by a linear differential equation $\dot{x} = Ax$ with system matrix $A \in \mathrm{Mat}(n, \mathbb{R})$, namely

$$\Phi_t : \mathbb{R}^n \to \mathbb{R}^n \quad , \quad x \mapsto \exp(At)x \qquad (t \in \mathbb{R}),$$

but we wish to obtain a deeper geometric insight. In particular, we will study the phase portraits of Φ for small dimensions n.
Generally speaking, the *phase portrait* of a dynamical system $\Phi : G \times M \to M$ refers to the decomposition of the phase space M into orbits.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_5

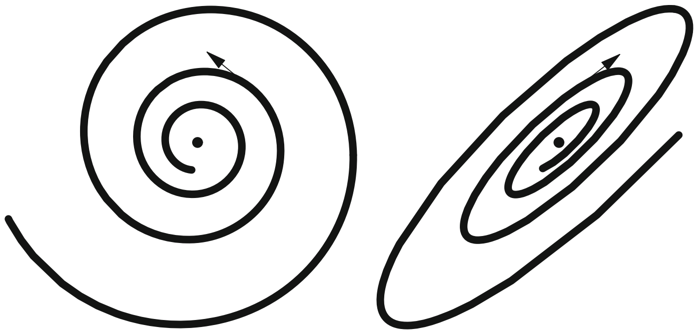


Figure 5.1.1 Phase portraits of stable spirals of the differential equation $\dot{x} = Ax$. Left: $A = \begin{pmatrix} -1/5 & -1 \\ 1 & -1/5 \end{pmatrix}$; right: a matrix that is similar to A

5.1 Conjugacies of Linear Flows

First let us ask, somewhat imprecisely, when a second linear differential equation $\dot{x} = Bx$ on the phase space $\mathbb{R}^n$ has a similar phase portrait as the one of $\dot{x} = Ax$. Maybe the following classification seems like a plausible start.

The matrices $A, B \in \mathrm{Mat}(n, \mathbb{R})$ are called *similar*, if there exists $S \in \mathrm{GL}(n, \mathbb{R})$ such that $B = SAS^{-1}$. In this case, the flow $\Psi_t(y) := \exp(Bt)y$ generated by B satisfies

$$\Psi_t(y) = S\exp(At)S^{-1}y = S\Phi_t(S^{-1}y)\,. \tag{5.1.1}$$

So the phase portrait for B arises from the one for A by a basis transformation of $\mathbb{R}^n$. In Figure 5.1.1, we see phase portraits of two similar matrices.

Since similarity transformations leave the eigenvalues and their multiplicities invariant, this classification is too fine for many purposes. What is more appropriate for comparing two continuous dynamical systems $\Phi^{(i)} : \mathbb{R} \times M^{(i)} \to M^{(i)}$ is the notion of a conjugacy by a homeomorphism $h \in C\big(M^{(1)}, M^{(2)}\big)$ (see Definition 2.28).

If the dynamical systems are differentiable and if h is even a *diffeo*morphism—i.e., $h \in C^1\big(M^{(1)}, M^{(2)}\big)$ and $h^{-1} \in C^1\big(M^{(2)}, M^{(1)}\big)$—, then it follows from $\Phi_t^{(2)} \circ h = h \circ \Phi_t^{(1)}$ that the vector fields f_k satisfy:

$$f_2 \circ h = \frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi_t^{(2)} \circ h\right)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\left(h \circ \Phi_t^{(1)}\right)\Big|_{t=0} = \mathrm{D}h \circ f_1\,,$$

or

$$f_2 = \mathrm{D}h \circ f_1 \circ h^{-1}\,. \tag{5.1.2}$$

So the vector fields generating the flows are mapped into each other by the linearization of h.

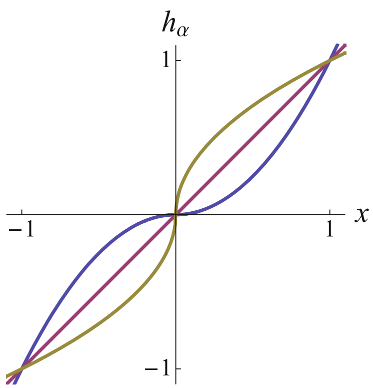
If in particular $x_1 \in M^{(1)}$ is an equilibrium of $\Phi^{(1)}$, then according to Exercise 2.30, $x_2 = h(x_1)$ will be an equilibrium of $\Phi^{(2)}$, and the linearizations $\mathrm{D}f_1(x_1)$ and $\mathrm{D}f_2(x_2)$ are similar matrices from $\mathrm{Mat}(n, \mathbb{R})$.

So if we apply this to the equilibrium $0 \in \mathbb{R}^n$, it follows for *linear* flows on $\mathbb{R}^n$ that such flows are conjugate by diffeomorphisms if and only if their vector fields are defined by similar matrices. In this case, we may however take the *linear* mapping defined in (5.1.1) by $S \in \mathrm{GL}(n, \mathbb{R})$ as a conjugating diffeomorphism, rather than having to take *general* diffeomorphisms of $\mathbb{R}^n$.

The situation is different when we use homeomorphisms h that need not be differentiable.

5.1 Example (Linear differential equations on $\mathbb{R}$) For a parameter $a \in \mathbb{R}$, we study the linear differential equation $\dot{x} = ax$ on $\mathbb{R}$, with its flow $\Phi_t^{(a)}(x) = e^{at}x$. Then the origin $x = 0$ is an equilibrium for all $a \in \mathbb{R}$. Now if $x \in \mathbb{R}\backslash\{0\}$, then the α- and ω-limit sets of x (see Definition 2.20) depend on the parameter:

- for $a < 0$: $\alpha(x) = \emptyset,\ \omega(x) = \{0\}$;
- for $a = 0$: $\alpha(x) = \omega(x) = \{x\}$;
- for $a > 0$: $\alpha(x) = \{0\},\ \omega(x) = \emptyset$.



The conjugating homeomorphism h_α, for $\alpha = 1/2,\ 1$ and 2

Due to Exercise 2.30, the flows $\Phi^{(a)}$ and $\Phi^{(b)}$ can only be conjugate if $\mathrm{sign}(a) = \mathrm{sign}(b)$. And in this case, the flows are indeed conjugate. For if a and b are both positive, or both negative, we can use the homeomorphism

$$h_\alpha : \mathbb{R} \to \mathbb{R},\, x \mapsto \mathrm{sign}(x)|x|^\alpha$$

of $\mathbb{R}$ (where $\alpha > 0$). Indeed, $h_\alpha^{-1} = h_{1/\alpha}$ and therefore choosing $\alpha := \frac{b}{a} > 0$, we have

$$\begin{aligned} h_\alpha \circ \Phi_t^{(a)} \circ h_\alpha^{-1}(x) &= h_\alpha(e^{at}\mathrm{sign}(x)|x|^{1/\alpha}) \\ &= \mathrm{sign}(x)(e^{at}|x|^{1/\alpha})^\alpha = e^{bt}\mathrm{sign}(x)|x| = e^{bt}x = \Phi_t^{(b)}(x). \end{aligned}$$

Note that this conjugating homeomorphism is smooth except at 0.

Thus the parameter space of one dimensional linear dynamical systems $\dot{x} = ax$ (which is itself one dimensional as it is parametrized by $a \in \mathbb{R}$) is decomposed into three equivalence classes by conjugacy. $\Diamond$

5.2 Hyperbolic Linear Vector Fields

We now generalize Example 5.1 to arbitrary dimensions.

5.2 Definition

- *A matrix* $A \in \mathrm{Mat}(n, \mathbb{R})$*, the vector field* $x \mapsto Ax$*, and the flow* $(t, x) \mapsto \Phi_t(x) = \exp(At)x$ *are called* **hyperbolic***, if all eigenvalues* $\lambda \in \mathbb{C}$ *of* A *satisfy* $\mathrm{Re}(\lambda) \neq 0$.
- *The sum of the algebraic multiplicities of those eigenvalues* λ *that satisfy* $\mathrm{Re}(\lambda) < 0$ *is called the* **index** *of* A *and is written as* $\mathrm{Ind}(A)$.
- $E^s := \{x \in \mathbb{R}^n \mid \lim_{t\to\infty} \Phi_t(x) = 0\}$ *is called the* **stable subspace***,* $E^u := \{x \in \mathbb{R}^n \mid \lim_{t\to-\infty} \Phi_t(x) = 0\}$ *the* **unstable subspace** *of* A.

5.3 Theorem *For all* $n \in \mathbb{N}$*, the set of hyperbolic matrices in* $\mathrm{Mat}(n, \mathbb{R})$ *is open and dense.*

Proof: • Let $A \in \mathrm{Mat}(n, \mathbb{R})$ be hyperbolic and let $\lambda \in \mathrm{i}\,\mathbb{R}$. Then λ is not an eigenvalue of A. The modulus of the characteristic polynomial diverges with the modulus of its argument: $\det(\lambda\, 1\!\!1_n - A) = \lambda^n\big(1 + \mathcal{O}(\|A\|/\lambda)\big)$, and the mapping $\det : \mathrm{Mat}(n, \mathbb{C}) \to \mathbb{C}$ is continuous. So

$$I(A) := \inf\{|\det(\lambda\, 1\!\!1_n - A)| \mid \lambda \in \mathrm{i}\,\mathbb{R}\} \;>\; 0\,,$$

and there exist $\Lambda > 0$ and a neighborhood $U \subset \mathrm{Mat}(n, \mathbb{R})$ of A with

$$I(B) = \inf\{|\det(\lambda\, 1\!\!1_n - B)| \mid \lambda \in \mathrm{i}\,\mathbb{R},\ |\lambda| \le \Lambda\} > 0 \qquad (B \in U).$$

So the matrices $B \in U$ are also hyperbolic.
• Assume $A \in \mathrm{Mat}(n, \mathbb{R})$ is not hyperbolic. Then for $c \in \mathbb{R}$, the matrix $A + c 1\!\!1_n \in \mathrm{Mat}(n, \mathbb{R})$ is hyperbolic provided $|c| \in (0, C)$ with

$$C := \inf\big\{|\mathrm{Re}(\lambda)| \mid \lambda \in \mathbb{C} \text{ eigenvalue of } A,\ \mathrm{Re}(\lambda) \neq 0\big\} \in (0, \infty]\,.$$

This set of matrices has A as an accumulation point. □

5.4 Remarks (Hyperbolic Matrices and Indices)

1. Whereas typical matrices in $\mathrm{Mat}(n, \mathbb{R})$ are hyperbolic, the same is no longer true when we restrict the discussion to the subspace of the infinitesimally symplectic matrices (see page 104), which are the ones that occur as system matrices in classical mechanics.
2. A matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ is hyperbolic if and only if

$$\mathrm{Ind}(A) + \mathrm{Ind}(-A) = n\,.$$

3. In Ex. 5.1, the hyperbolic dynamical systems with the same index were conjugate to each other. We will now prove the same for arbitrary dimensions n.
4. The (commonly used) indices s and u refer to *stable* and *unstable* respectively.
5. The *index of an equilibrium* x_0 of a dynamical system $\dot{x} = f(x)$ is defined to be the index of $\mathrm{D}f(x_0)$. For a linear vector field ($f(x) = Ax$), the index of any equilibrium, in particular 0, is therefore the index of the system matrix A. ◊

5.5 Exercise (Index) Determine a fundamental system of solutions to

$$\dot{x} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & -1 & 2 \\ -2 & -1 & 1 \end{pmatrix} x.$$

Which solutions remain bounded as $t \to \infty$? Is the matrix hyperbolic, and if so, what is its index? ◊

5.6 Lemma (Index) *For a hyperbolic matrix A, one has* $\dim(E^s) = \mathrm{Ind}(A)$.

Proof: First note that, due to the linearity of the flow Φ_t, the sets E^u and E^s are indeed subspaces of the phase space $\mathbb{R}^n$.

- If $x \in \mathbb{R}^n$ is an element of the direct sum of the generalized eigenspaces for the eigenvalues λ_i that satisfy $\mathrm{Re}(\lambda_i) < 0$, then the components of the vector valued function $t \mapsto \Phi_t(x)$ are sums of λ_i-quasipolynomials, hence $x \in E^s$. This shows $\dim(E^s) \geq \mathrm{Ind}(A)$.
- On the other hand, a similar argument for the eigenvalues with positive real part shows that $\dim(E^u) \geq \mathrm{Ind}(-A) = n - \mathrm{Ind}(A)$.
- Furthermore, for any sum $f(t) := \sum_i p_i(t)e^{\lambda_i t}$ of λ_i-quasipolynomials, one has: If $\lim_{t\to\infty} f(t) = \lim_{t\to-\infty} f(t) = 0$, then $f = 0$.
  So we have $E^u \cap E^s = \{0\}$ and thus $\dim(E^u) + \dim(E^s) = n$, which now implies $\dim(E^s) = \mathrm{Ind}(A)$ and $\dim(E^u) = n - \mathrm{Ind}(A)$. □

5.7 Example The figure at the beginning of this chapter (page 79) shows orbits of a linear flow on $\mathbb{R}^3$ with index 3. ◊

If we look at the phase portrait on the right in Figure 5.1.1, we see that the trajectory is stable, but does not approach the origin monotonically. This deficiency can be mended by transition to the similar system matrix on the left side of the figure. Such a transition is possible in generality:

5.8 Lemma *Let* $A \in \mathrm{Mat}(n, \mathbb{R})$ *and* $\Lambda := \max\{\mathrm{Re}(\lambda) \mid \lambda$ *eigenvalue of* $A\}$. *Then, for each* $\Lambda' > \Lambda$ *there exists a scalar product on* $\mathbb{R}^n$ *whose associated norm allows the estimate:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\Phi_t(x)\| \leq \Lambda' \,\|\Phi_t(x)\| \qquad (x \in \mathbb{R}^n,\ t \in \mathbb{R}). \tag{5.2.1}$$

Proof: Because of $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t(x) = \frac{\mathrm{d}}{\mathrm{d}s}\Phi_{t+s}(x)|_{s=0} = \frac{\mathrm{d}}{\mathrm{d}s}\Phi_s(y)|_{s=0}$ with $y := \Phi_t(x)$, it suffices to show (5.2.1) for $t = 0$.

- Instead of (5.2.1), we show for an appropriate scalar product on $\mathbb{C}^n$ that

$$\tfrac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle \exp(At)x,\ \exp(At)x\rangle\,|_{t=0} \le \Lambda'\,\langle x, x\rangle \qquad (x \in \mathbb{C}^n). \qquad (5.2.2)$$

  Without loss of generality, we may assume that the change of basis to Jordan normal form has already been performed. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \exp(At)x,\ \exp(At)x\rangle\,|_{t=0} = 2\mathrm{Re}\big(\langle x,\ Ax\rangle\big)\,.$$

- If the scalar product is chosen in such a way that subspaces corresponding to different Jordan blocks are orthogonal, this expression is a sum over the contributions from each Jordan block.
  If $J_r(\lambda)$ is a Jordan block for eigenvalue λ, then for $\mu := \mathrm{Re}(\lambda) \le \Lambda$, we get

$$\mathrm{Re}\big(\langle x,\ J_r(\lambda)x\rangle\big) = \mathrm{Re}\big(\langle x,\ J_r(\mu)x\rangle\big)\,.$$

- For $\varepsilon > 0$, we conjugate $J_r(\mu) = \mu 1\!\!1_r + J_r(0)$ with the diagonal matrix $D_\varepsilon := \mathrm{diag}(1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{r-1}) \in \mathrm{GL}(r, \mathbb{C})$ (hence $D_\varepsilon^{-1} = D_{1/\varepsilon}$):

$$D_\varepsilon^{-1} J_r(\mu) D_\varepsilon = \mu 1\!\!1_r + D_\varepsilon^{-1} J_r(0) D_\varepsilon = \mu 1\!\!1_r + \varepsilon J_r(0)\,.$$

  So the off-diagonal has been multiplied by ε. Let us denote the canonical scalar product on $\mathbb{C}^r$ by $\langle\cdot,\cdot\rangle_{\mathrm{can}}$. The Cauchy-Schwarz inequality implies

$$\mathrm{Re}\big(\langle x,\ J_r(0)x\rangle_{\mathrm{can}}\big) \le \|x\|_{\mathrm{can}}\,\|J_r(0)x\|_{\mathrm{can}} \le \|x\|_{\mathrm{can}}^2\,.$$

  Therefore, for $\varepsilon \in \big(0, \Lambda' - \Lambda\big)$, we have

$$\begin{aligned}\mathrm{Re}\big(\langle x,\ (\mu 1\!\!1_r + \varepsilon J_r(0))x\rangle_{\mathrm{can}}\big) &= \mu\,\langle x,\ x\rangle_{\mathrm{can}} + \varepsilon\,\mathrm{Re}\big(\langle x,\ J_r(0)x\rangle_{\mathrm{can}}\big)\\ &\le (\Lambda+\varepsilon)\,\langle x,\ x\rangle_{\mathrm{can}} \le \Lambda'\,\langle x,\ x\rangle_{\mathrm{can}}\ .\end{aligned}$$

  So we define the scalar product by $\langle x,\ y\rangle := \big\langle D_\varepsilon^{-1}x,\ D_\varepsilon^{-1}y\big\rangle_{\mathrm{can}}$ and obtain for $\tilde{x} := D_\varepsilon^{-1}x$:

$$\mathrm{Re}\,\langle x, J_r(\mu)x\rangle = \mathrm{Re}\,\langle \tilde{x},\ (\mu 1\!\!1_r + \varepsilon J_r(0))\tilde{x}\rangle_{\mathrm{can}} \le \Lambda'\,\langle \tilde{x},\ \tilde{x}\rangle_{\mathrm{can}} = \Lambda'\,\langle x,\ x\rangle\,,$$

  hence altogether $\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\exp(At)x\|^2|_{t=0} \le \Lambda'\|x\|^2$, i.e., (5.2.2). □

5.9 Theorem (Conjugacy Classes)
The linear flows of two hyperbolic matrices $A^{(1)}, A^{(2)} \in \mathrm{Mat}(n, \mathbb{R})$ *are conjugate if and only if*

$$\mathrm{Ind}\big(A^{(1)}\big) = \mathrm{Ind}\big(A^{(2)}\big)\,.$$

Proof: We denote the linear flows as $\Phi_t^{(i)}(x) := \exp(A^{(i)}t)x$, and their stable subspaces as $E^{(i)}$ $(i = 1, 2)$. Let $\|\cdot\|^{(i)} : E^{(i)} \to [0, \infty)$ be norms that satisfy the inequality from Lemma 5.8 for some $\Lambda < 0$.

- If a conjugating homeomorphism $h : \mathbb{R}^n \to \mathbb{R}^n$ exists, then $h(0) = 0$ using Exercise 2.30 (a). In addition $h\big(E^{(1)}\big) = E^{(2)}$, because the points in $x \in E^{(i)}$ can be defined by the characteristic property that $\{0\} = \omega(x)$, and Exercise 2.30 (c) provides $h\big(\omega(x)\big) = \omega\big(h(x)\big)$. An important property of homeomorphisms of vector spaces is that they leave the dimension invariant.[1] This implies that, if a conjugacy h exists, then

$$\mathrm{Ind}\big(A^{(1)}\big) = \dim\big(E^{(1)}\big) = \dim\big(h(E^{(1)})\big) = \dim\big(E^{(2)}\big) = \mathrm{Ind}\big(A^{(2)}\big).$$

- We now assume conversely that $\mathrm{Ind}\big(A^{(1)}\big) = \mathrm{Ind}\big(A^{(2)}\big)$ and construct a conjugating homeomorphism h. As each of the two phase spaces $\mathbb{R}^n$ is the direct sum of its stable and unstable subspaces, we write the homeomorphism as $h = \big(h^{(s)}, h^{(u)}\big)$ with respect to these decompositions, so that $h^{(s)} : E^{(1)} \to E^{(2)}$, and $h^{(u)}$ similarly maps the unstable subspaces to each other.
- Now we define $h^{(s)}$ and show that this mapping is a homeomorphism. First, $h^{(s)}(0) := 0$, because equilibria are unique here, and are mapped to equilibria. Next, if $x \in E^{(1)}\backslash\{0\}$, then $\lim_{t\to+\infty} \Phi_t^{(1)}(x) = 0$ and $\lim_{t\to-\infty} \big\|\Phi_t^{(1)}(x)\big\|^{(1)} = \infty$. By 5.8, there is exactly one

$$T(x) \in \mathbb{R} \quad \text{with} \quad \Phi_{T(x)}^{(1)}(x) \in S^{(1)} := \Big\{y \in E^{(1)} \;\Big|\; \|y\|^{(1)} = 1\Big\};$$

  and $T : E^{(1)} \setminus \{0\} \to \mathbb{R}$ is smooth by the implicit function theorem and the smoothness of the flow. In geometric terms, $S^{(1)}$ is the unit sphere in $E^{(1)}$.
  Similarly, $S^{(2)} := \{z \in E^{(2)} \mid \|z\|^{(2)} = 1\}$ will denote the unit sphere in $E^{(2)}$. As the two $\mathbb{R}$-vector spaces $E^{(1)}$ and $E^{(2)}$ have the same dimension, there is an isomorphism $I : E^{(1)} \to E^{(2)}$, and correspondingly, the diffeomorphism

$$\tilde{I} : S^{(1)} \to S^{(2)} \quad , \quad y \mapsto \frac{I(y)}{\|I(y)\|^{(2)}}.$$

  We set

$$h^{(s)}(x) := \Phi_{-T(x)}^{(2)} \circ \tilde{I} \circ \Phi_{T(x)}^{(1)}(x) \qquad \big(x \in E^{(1)}\backslash\{0\}\big).$$

  Being a composition of smooth maps, $h^{(s)}$ is smooth on $E^{(1)}\backslash\{0\}$, and $h^{(s)}(x) \to 0$ for $x \to 0$. The analogous statements hold for the inverse mapping of $h^{(s)}$. This shows that $h^{(s)}$ is a homeomorphism.
  $h^{(u)}$ is defined analogously, by working with the system matrices— $A^{(i)}$ instead of $A^{(i)}$. Thus, finally, h is a homeomorphism, too.

---

[1]This is a nontrivial result, using tools from algebraic topology.

- h is a conjugacy between the flows. To see this, note first that for $y := \Phi_s^{(1)}(x)$ and $x \in E^{(1)} \setminus \{0\}$, one has

$$\Phi^{(1)}_{T(x)-s}(y) = \Phi^{(1)}_{T(x)} \circ \Phi^{(1)}_{-s}(y) = \Phi^{(1)}_{T(x)}(x)\,, \quad \text{hence } T(y) = T(x) - s\,.$$

  Next, for all $t \in \mathbb{R}$ and (initially, only) for $x \in E^{(1)}$, one finds therefore that $\Phi_t^{(2)} \circ h(x)$ equals

$$\Phi^{(2)}_{t-T(x)} \circ \tilde{I} \circ \Phi^{(1)}_{T(x)}(x) = \Phi^{(2)}_{t-T(x)} \circ \tilde{I} \circ \Phi^{(1)}_{T(x)-t} \circ \Phi^{(1)}_t(x) = h \circ \Phi^{(1)}_t(x)\,.$$

  The analogous argument for the unstable subspace implies the conjugacy property for all $x \in \mathbb{R}^n$. □

We have thus found that under conjugacy, there are exactly $n+1$ equivalence classes of hyperbolic matrices $A \in \mathrm{Mat}(n, \mathbb{R})$.

5.10 Literature A more extensive analysis, in particular a generalization to the local theory of nonlinear differential equations near a hyperbolic singular point, can be found in PALIS and DE MELO [PdM], and also in AMANN [Am]. ◊

5.3 Linear Flows in the Plane

After the case of phase space dimension $n = 1$, which was discussed in Example 5.1, we now study the next simplest case $n = 2$.

So we study for $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathrm{Mat}(2, \mathbb{R})$ the flow $\Phi^{(A)} : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ of the linear ODE $\dot{x} = Ax$. Similar matrices lead to flows that differ only by a basis transformation. The quantities

$$\mathrm{tr}(A) = a_{11} + a_{22}\,, \ \det(A) = a_{11}a_{22} - a_{12}a_{21} \text{ and } D(A) := \mathrm{tr}(A)^2 - 4\det(A)$$

are invariant under conjugations $A \mapsto SAS^{-1}$, and the eigenvalues $\lambda_{1/2} \in \mathbb{C}$ equal

$$\lambda_{1/2} = \tfrac{1}{2}\left(\mathrm{tr}(A) \pm \sqrt{D}\right).$$

It is only for discriminant $D = 0$ that the complex Jordan normal form of A can have a Jordan block of size 2, and it is only in that case of a double eigenvalue that the similarity class of A will not already be determined by $\mathrm{tr}(A)$ and $\det(A)$.

The matrix A fails to be hyperbolic if and only if at least one of the eigenvalues has vanishing real part. This happens exactly when

1. $\det(A) = 0$, i.e., one eigenvalue is actually 0, or
2. $\det(A) > 0$, but $\mathrm{tr}(A) = 0$, i.e., the eigenvalues are purely imaginary.

These conditions form the horizontal and the positive vertical axis in the plane of $(\mathrm{tr}, \det) \in \mathbb{R}^2$, so they separate three domains,[2] see Figure 5.3.1.

- $\mathrm{Ind}(A) = 0$ applies to the quadrant with $\det(A) > 0$ and $\mathrm{tr}(A) > 0$.
- $\mathrm{Ind}(A) = 1$ applies when $\det(A) < 0$. In this case, both eigenvalues are real.
- $\mathrm{Ind}(A) = 2$ corresponds to the quadrant with $\det(A) > 0$ and $\mathrm{tr}(A) < 0$.

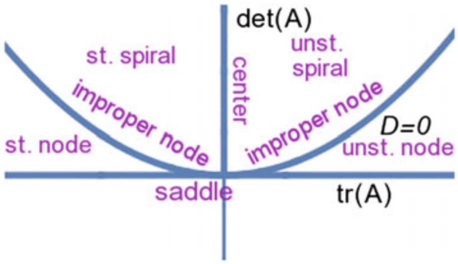


Figure 5.3.1 Bifurcation diagram for matrices $A \in \mathrm{Mat}(2, \mathbb{R})$, with discriminant $D \equiv D(A) = \mathrm{tr}(A)^2 - 4\det(A)$. $\mathrm{tr}(A) < 0$: sink; $\mathrm{tr}(A) = 0$: volume preserving flow; $\mathrm{tr}(A) > 0$: Source

As shown in the previous section, flows inside each of these three domains are conjugate to each other, but flows of matrices with different index are not conjugate.

The case $\mathrm{Ind}(A) = 1$ can be further subdivided, according to the sign of $\mathrm{tr}(A)$. In the case $\mathrm{tr}(A) = 0$ with two eigenvalues $\lambda_1 = -\lambda_2 \in \mathbb{R}$, the phase space volume is conserved under the flow, whereas for $\mathrm{tr}(A) < 0$ it tends to 0 with an exponential rate according to Lemma 4.12. The case $\mathrm{tr}(A) > 0$ is shown in Figure 5.3.2.

The parabola defined by the equation $D = 0$ further subdivides the conjugacy classes $\mathrm{Ind}(A) = 0$ and $\mathrm{Ind}(A) = 2$. For $D > 0$, i.e., real eigenvalues, we obtain phase space portraits called *nodes*, see Figure 5.3.3. These nodes are called *stable* when $\dim(E^s) = 2$, i.e., $\mathrm{Ind}(A) = 2$ and *unstable* when $\mathrm{Ind}(A) = 0$.

For $D = 0$, the Jordan normal form of A could consist of a single nontrivial Jordan block. A corresponding phase portrait, called *improper node*, can also be found in Figure 5.3.3.

If in addition, $\mathrm{tr}(A) = 0$, i.e., both eigenvalues vanish, one obtains a one-dimensional eigenspace of equilibria, as in Figure 5.3.4 (left). The case $\mathrm{tr}(A) = 0$, $\det(A) > 0$ leads to imaginary eigenvalues and periodic orbits (Figure 5.3.4 right), these are also called *centers*.

Finally, for $D < 0$ and $\mathrm{tr}(A) < 0$, the motion is described by spirals, and is stable (Figure 5.3.5), whereas $D < 0$ and $\mathrm{tr}(A) > 0$ leads to *unstable spirals*.

The (Hamiltonian) case of classical mechanics without friction corresponds to a matrix $A \in \mathrm{Mat}(2, \mathbb{R})$ with $\mathrm{tr}(A) = 0$. So we are on the vertical axis of the bifurcation diagram 5.3.1.

[2] A **domain** is defined as an open, nonempty and connected subset of a topological space.

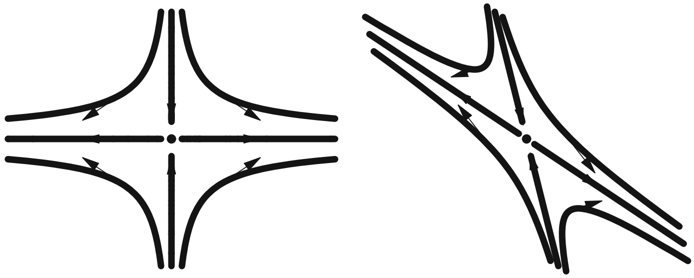


Figure 5.3.2 Phase portrait of saddles for the differential equation $\dot{x} = Ax$. Left: system matrix $A = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$; Right: a matrix similar to A

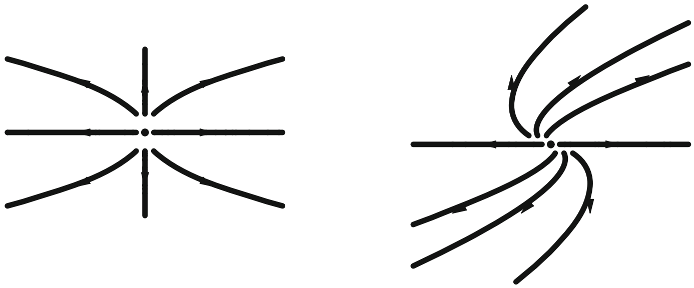


Figure 5.3.3 Phase portraits of nodes for the ODE $\dot{x} = Ax$. Left: unstable node, for $A = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$; Right: unstable improper node, for $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

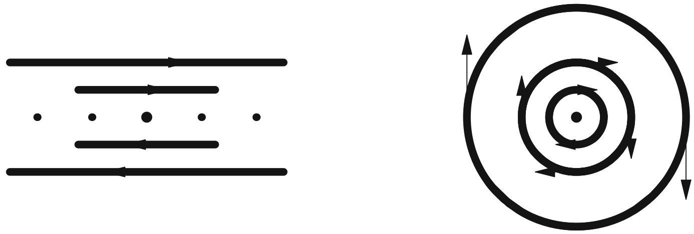


Figure 5.3.4 Phase portraits for $\dot{x} = Ax$: the case of purely imaginary eigenvalues. Left: the nilpotent matrix $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$; Right: Center, for the antisymmetric matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

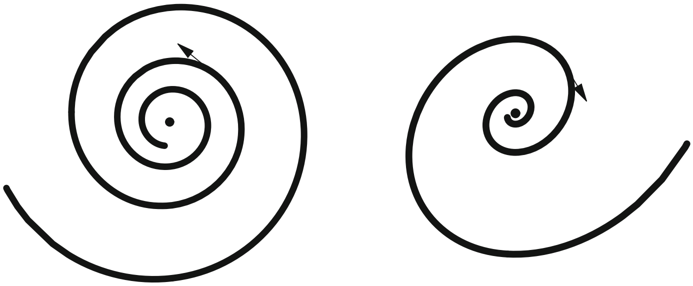


Figure 5.3.5 Phase portraits of stable spirals for the differential equation $\dot{x} = Ax$. Left: $A = \left(\begin{smallmatrix} -1/5 & -1 \\ 1 & -1/5 \end{smallmatrix}\right)$; right: a matrix that is not similar to A

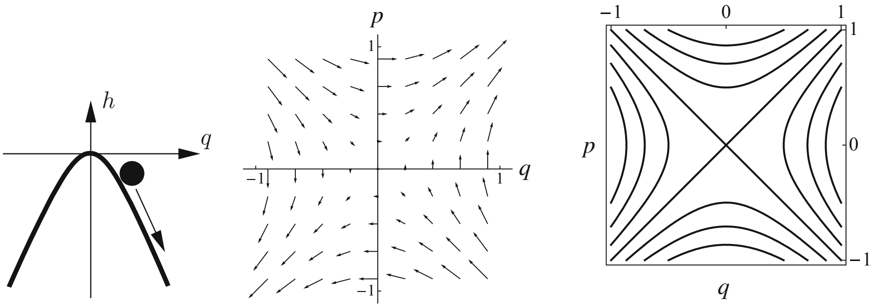


Figure 5.3.6 Repelling force ($a = 1$): vector field $\dot{q} = p$, $\dot{p} = q$ and its phase portrait

5.11 Example (Hamiltonian Linear Differential Equations)
We consider a mass point of mass 1 at location $q \in \mathbb{R}$ that is accelerated by a force $F(q) := a\,q$, with a parameter $a \in \mathbb{R}$. In terms of physics, this could represent an object that slides under the influence of gravitation and without friction on a parabola shaped surface near its minimum (see also Exercise 8.21 for the true solution). According to Newton, the second order ODE $\ddot{q} = a\,q$ applies. By introducing the velocity $p = \dot{q}$ one obtains the linear ODE system of first order

$$\dot{q} = p \quad , \quad \dot{p} = a\,q$$

or $\dot{x} = A\,x$ with $x = \begin{pmatrix} q \\ p \end{pmatrix}$ and $A := \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix}$. ◊

5.12 Exercise (Hooke's Law) Show that the linear flow

$$\big(q(t), p(t)\big) = \Phi_t\big((q_0, p_0)\big) = \exp(A\,t)\begin{pmatrix} q_0 \\ p_0 \end{pmatrix} \quad \text{for} \quad A = \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix}$$

satisfies the following:

(a) If $a > 0$ (see Figure 5.3.6), then

$$\big(q(t), p(t)\big) = \left(q_0 \cosh(\omega t) + \frac{p_0}{\omega} \sinh(\omega t)\ ,\ \omega q_0 \sinh(\omega t) + p_0 \cosh(\omega t)\right),$$

where $\omega := \sqrt{a}$.

(b) If $a = 0$ (see Figure 5.3.7), then

$$\big(q(t), p(t)\big) = (q_0 + p_0 t, p_0).$$

(c) If $a < 0$, (see Figure 5.3.8), then

$$\big(q(t), p(t)\big) = \left(q_0 \cos(\omega t) + \frac{p_0}{\omega} \sin(\omega t)\ ,\ -\omega q_0 \sin(\omega t) + p_0 \cos(\omega t)\right),$$

with $\omega := \sqrt{-a}$. ◊

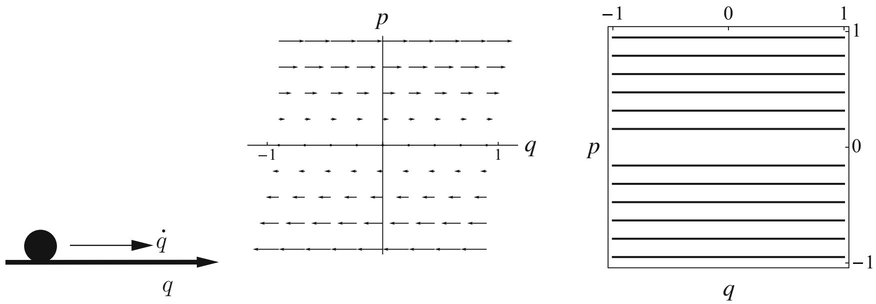


Figure 5.3.7 Free motion ($a = 0$): the vector field $\dot{q} = p$, $\dot{p} = 0$ and its phase portrait

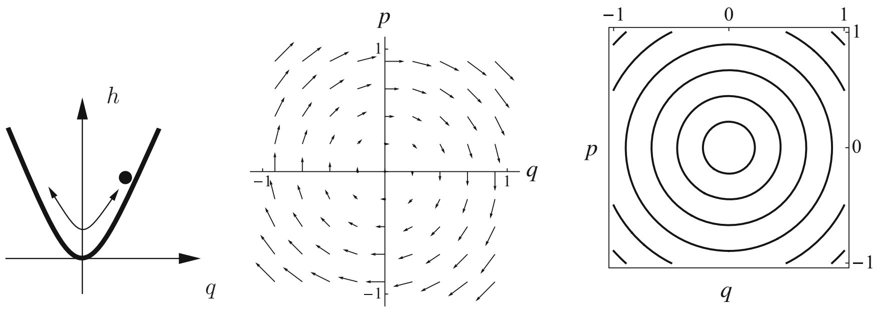


Figure 5.3.8 Attracting force ($a = -1$): the vector field $\dot{q} = p$, $\dot{p} = -q$ and its phase portrait

We note that in all three cases, $\det(\exp(At)) = 1$ for $t \in \mathbb{R}$. This also follows from Theorem 4.12. An intuitive interpretation of this fact is the observation that the flow in $\mathbb{R}^2$ is area preserving.

5.4 Example: Spring with Friction

As an application of the theory of linear differential equations, we discuss the example of a mass $m > 0$ suspended from a spring; its equilibrium is to be at height $x = 0$.

The force $F(x, \dot{x})$ that acts on the mass is, in the simplest approximation, a linear function of the elongation $x \in \mathbb{R}$ and the velocity $\dot{x} \in \mathbb{R}$, that is, $F(x, \dot{x}) = -Dx - R\dot{x}$. The first constant of proportionality, $D > 0$, is called the stiffness constant because it is a measure for the stiffness of the spring.

The second constant, $R \geq 0$ describes the friction of the massive body in the surrounding air as well as internal friction of the material that makes up the spring.[3]

Autonomous Case

So by Newton's law, we have $m\frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t) = -Dx(t) - R\frac{\mathrm{d}}{\mathrm{d}t}x(t)$. With the new time parameter $s = \sqrt{\frac{m}{D}}t$ and the abbreviation $\dot{x}$ for $\frac{\mathrm{d}}{\mathrm{d}s}x(t(s))$, we get the linear homogeneous ODE

$$\ddot{x} = -x - k\dot{x} \quad , \text{with} \quad k := \frac{R}{\sqrt{mD}} \geq 0\,. \tag{5.4.1}$$

This kind of rescaling is commonly used to bring a differential equation to its simplest possible form.

Introducing the velocity $v := \dot{x}$, one obtains the linear system of first order

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = A \begin{pmatrix} x \\ v \end{pmatrix} \quad \text{with} \quad A := \begin{pmatrix} 0 & 1 \\ -1 & -k \end{pmatrix}.$$

The eigenvalues of A are obtained as the zeros $\lambda_{1/2} = -\frac{k}{2} \pm \mathrm{i}\sqrt{1 - \frac{k^2}{4}}$ of the characteristic polynomial $\det(\lambda\mathbb{1} - A) = \lambda^2 + k\lambda + 1$.

As $\det(A) = 1$ and $\mathrm{tr}(A) = -k$, we find our system along a horizontal line in the diagram 5.3.1. Depending on the magnitude of the friction term, three cases need to be distinguished:

1. **Oscillatory case:** Small friction, $0 \leq k < 2$, (see also Exercise 5.12).

   In this case, the general solution is of the form

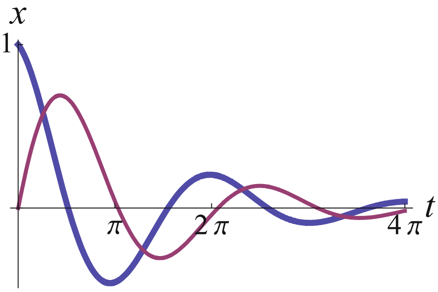
$$x(t) = e^{-kt/2}\big(a\cos(\omega(k)t) + b\sin(\omega(k)t)\big)\,,$$

[3] In contrast to this friction, which is proportional to the velocity and named after *Stokes*, the *friction* within a turbulent fluid is empirically described as being proportional to the square of the velocity.

where $\omega(k) := \mathrm{Im}(\lambda_1) = \sqrt{1 - \frac{k^2}{4}}$, and the coefficients a and b are to be determined from the initial values $x(0)$, $\dot{x}(0)$.

In the frictionless case $k = 0$, we have a center, otherwise a spiral. The frequency $\omega(k)$ of the oscillation is reduced, compared to $\omega(0) = 1$, but it is still true that $\omega(k) > 0$.

The mass suspended from a spring will gradually oscillate towards its equilibrium $(x, \dot{x}) = (0, 0)$.



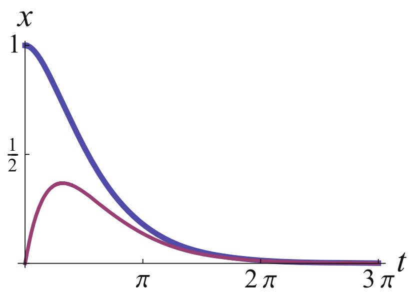
Two solutions for the oscillatory case ($k = 1/2$)

2. **Critically damped case:** $k = 2$.

In this case, $A = \begin{pmatrix} 0 & 1 \\ -1 & -2 \end{pmatrix}$, and the matrix $V := \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ with $V^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ transforms A into upper triangular form:

$$J := V^{-1}AV = \begin{pmatrix} -1 & 2 \\ 0 & -1 \end{pmatrix}.$$

Therefore $\exp(Jt) = e^{-t}\begin{pmatrix} 1 & 2t \\ 0 & 1 \end{pmatrix}$ and $\exp(At) = Ve^{Jt}V^{-1} = e^{-t}\begin{pmatrix} 1+t & t \\ -1 & 1-t \end{pmatrix}$. For instance when the initial velocity vanishes, $v_0 = 0$, the solution is



Two solutions for the critically damped case ($k = 2$)

$$x(t) = x_0(1+t)e^{-t}.$$

So there is no oscillation left. Due to the nontrivial Jordan block, the motion to the equilibrium is slowed down compared to the solution $x(t) = x_0 e^{-t}$.

3. **Overdamped case:** Large friction, $k > 2$.
   In this case, A has two negative real eigenvalues

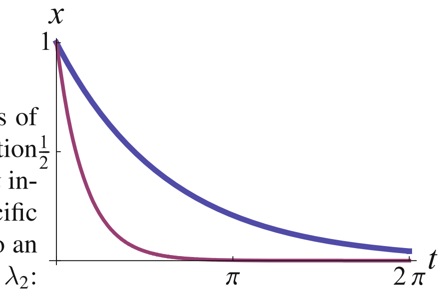
$$\lambda_{1/2} = -\frac{k \pm \sqrt{k^2 - 4}}{2} = -\frac{k}{2}\left(1 \pm \sqrt{1 - \frac{4}{k^2}}\right).$$

   For $k \to \infty$ one has

$$\lambda_1 \sim -k^{-1} \quad , \quad \lambda_2 \sim -k\,.$$

   The phase portrait is a node. In terms of physics, this means that the approximation to the equilibrium is slowed down as k increases, except in the case of very specific initial values x_0, v_0 that correspond to an eigenvector for the smaller eigenvalue λ_2:

$$x(t) = ae^{\lambda_1 t} + be^{\lambda_2 t}\,.$$



Two solutions for the overdamped case ($k = 2.5$)

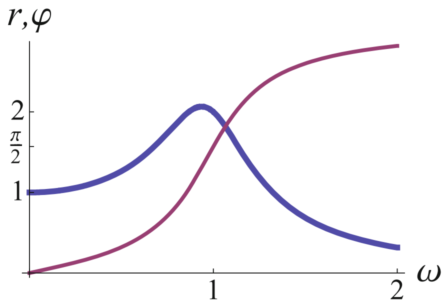
Non-autonomous case

A practically important generalization of the example just discussed is the situation when the mass point is under the influence of an external forcing. For instance, one could move the suspension point up and down periodically in time, in order to maintain a permanent oscillation. Then the differential equation under consideration has the following normal form:

$$\ddot{x} + k\dot{x} + x = f(t)\,,$$

where the given function f represents the external forcing, for example $f(t) = A\cos(\omega t)$. But as $\cos(\omega t) = \mathrm{Re}(e^{i\omega t})$, it is plausible to simplify the calculation by looking for a particular solution $y : \mathbb{R} \to \mathbb{C}$ to the complex differential equation

$$\ddot{y} + k\dot{y} + y = Ae^{i\omega t}$$



Amplitude and phase shift of the forced oscillation ($k = \frac{1}{2}$)

and then set $x(t) := \mathrm{Re}(y(t))$.

In terms of physics, one expects that, due to friction, the mass point will after some time mainly oscillate harmonically with the frequency[4] ω that is externally imposed on it. With the ansatz $y(t) := Be^{i\omega t}$, one obtains $y^{(j)}(t) = (i\omega)^j y(t)$, hence

$$(1 - \omega^2 + ik\omega)y(t) = Ae^{i\omega t}\,,$$

[4]called ‘circular frequency’ in physics.

or, solving for B:
$B = \frac{A}{1-\omega^2+ik\omega} = Ar_k(\omega)e^{-i\varphi_k(\omega)}$ with *amplitude* $r_k(\omega) := \frac{1}{\sqrt{(1-\omega^2)^2+(k\omega)^2}}$ and *phase* $\varphi_k(\omega) := \operatorname{arccot}\left(\frac{1-\omega^2}{k\omega}\right)$ of the forced oscillation (see the figure, which is called a *Bode diagram*).

The case $\omega = 1$ plays a special role; this comes as no surprise, because the homogeneous equation without friction had this same frequency in its solutions.

Let us interpret the amplitude and phase in terms of physics:

- **Amplitude:** For a small frequency ω, the mass oscillates in accordance with the amplitude of the forcing, since $r_k(0) = 1$.

  As ω gets close to the eigenfrequency (which for damping $k = 0$ was 1 by our normalization), *resonance* occurs. The amplitude of the oscillation will be enlarged in comparison with the forcing amplitude, and this effect is the larger, the smaller the damping.[5]
  The maximum of the amplitude $r_k(\omega)$ occurs when ω equals

  $$\omega_{\max}(k) := \sqrt{1 - k^2/2}\,.$$

  The maximal amplitude diverges as the damping goes to 0:

  $$r_k\big(\omega_{\max}(k)\big) = \frac{1}{k\sqrt{1-k^2/4}} \sim \frac{1}{k} \quad \text{as } k \searrow 0\,.$$

- **Phase:** The phase shift $\varphi_k(\omega) = \operatorname{arccot}\left(\frac{1-\omega^2}{k\omega}\right)$, by which the responding oscillation lags behind the forcing, has the following properties:
  - For $\omega \ll 1$, the external forcing and the resulting oscillation are in phase, namely $\lim_{\omega\searrow 0}\varphi_k(\omega) = 0$.
  - As $\lim_{\omega\nearrow+\infty}\varphi_k(\omega) = \pi$, they are in counter-phase for $\omega \gg 1$.
  - For $\omega = 1$, they have a phase shift of $\varphi_k(1) = \frac{\pi}{2}$. Then the external force and the velocity are in phase.

So we obtain the general solution of the inhomogeneous differential equation

$$\ddot{x} + k\dot{x} + x = A\cos(\omega t)$$

for $0 < k < \infty$ as

$$x = x_H + x_I \quad \text{with} \quad x_I(t) = r_k(\omega)A\cos\big(\omega t - \varphi_k(\omega)\big)\,,$$

and x_H a solution to the homogeneous problem (5.4.1).

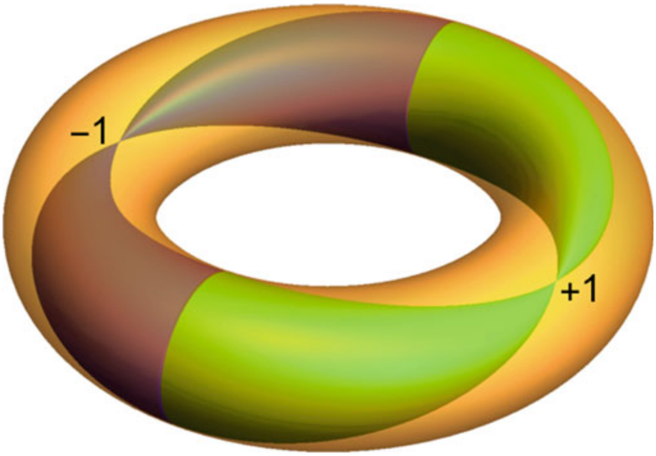
[5]We assume here a damping $k < \sqrt{2}$.

For $k = 0$, however, we get the particular solution

$$x_I(t) = \frac{A}{2} t \sin(t) \,,$$

a catastrophic resonance to be avoided in mechanical applications.

Chapter 6
Hamiltonian Equations and Symplectic Group



The symplectic group Sp(2,R) = SL(2,R) with the matrices ±1l and the hypersurface of matrices with degenerate eigenvalues +1 (right half) and -1 (left half) respectively.

The notion of energy may well be the most important notion in physics. In any case, the Hamilton function of the system (which is its total energy) defines the dynamics of the particles. The vector field of the differential equation arises by a rotation from the gradient of the Hamilton function. In the linear case, the dynamics happens within the symplectic group.

6.1 Gradient Flows and Hamiltonian Systems

In this chapter, we will deal with differential equations whose vector fields are defined in terms of a single function: Gradient systems and Hamiltonian differential

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_6

equations. These are frequently studied, and the latter are foundational for classical mechanics. On the other hand, their dynamics enjoys special properties.

6.1.1 Gradient Systems

We consider the *gradient system* of $H \in C^2(M, \mathbb{R})$ on the phase space M, which is an open subset of $\mathbb{R}^n$:

$$\begin{cases} \dot{x}_1 = \frac{\partial H}{\partial x_1}(x_1, \ldots, x_n) \\ \quad \vdots \\ \dot{x}_n = \frac{\partial H}{\partial x_n}(x_1, \ldots, x_n) \end{cases} \quad \text{or briefly } \dot{x} = \nabla H(x) \tag{6.1.1}$$

Here ∇ denotes the gradient with respect to the canonical metric on $\mathbb{R}^n$ (See page 503). If H satisfies the appropriate hypotheses, the differential equation above defines a differentiable dynamical system $\Phi : \mathbb{R} \times M \to M$, called the *gradient flow*.

We observe that H increases along orbits:

6.1 Lemma *Either the orbit of a gradient system (6.1.1) is an equilibrium, or else H is strictly increasing along the solution curve.*

Proof: Let $t \mapsto \varphi(t)$ be a solution curve, i.e., $\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t) = \nabla H\big(\varphi(t)\big)$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t} H\big(\varphi(t)\big) = \big\langle \nabla H\big(\varphi(t)\big), \dot{\varphi}(t)\big\rangle = \|\nabla H\big(\varphi(t)\big)\|^2 \geq 0\,.$$

If the gradient of H at $\varphi(t)$ is 0, then the orbit is an equilibrium. Otherwise, the strict inequality holds (for the entire orbit). □

The superlevel sets

$$M^c := \{m \in M \mid H(m) \geq c\} \qquad (c \in \mathbb{R})$$

are therefore forward invariant ($\Phi_t(M^c) \subseteq M^c \quad (t \geq 0)$), i.e., the trajectories are 'trapped' in M^c for all positive times. Since H increases along orbits, as we have seen, one gets the following

Conclusion:
Equation (6.1.1) does not have any periodic orbits other than equilibria.

6.2 Example $H : \mathbb{R}^2 \to \mathbb{R},\ H(x_1, x_2) := x_2^2 + (x_1^2 - 1)^2$.
The gradient flow exists for *all* negative times (but only some positive times) and has the equilibria $(0, 0)$, $(1, 0)$ and $(-1, 0)$, see Figure 6.1.1. ◊

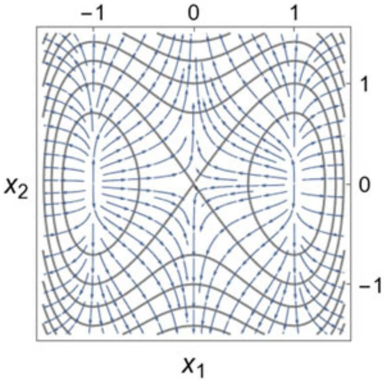


Figure 6.1.1 Level sets of $H(x) = x_2^2 + (x_1^2 - 1)^2$, and gradient flow of ∇H

6.1.1.1 Tests, Properties, Examples

How to decide whether a given vector field f is a gradient vector field? The theory of differential forms is helpful (Appendix B). For in cartesian coordinates, ∇H is simply the transpose of the exact 1-form $dH = \frac{\partial H}{\partial x_1} dx_1 + \cdots + \frac{\partial H}{\partial x_n} dx_n$. Therefore, it is necessary that the 1-form

$$\omega := f_1 dx_1 + \cdots + f_n dx_n$$

associated with the vector field $f = (f_1, \ldots, f_n)^\top$ be *closed* ($d\omega = 0$) if f is to be a gradient vector field.
For example in the case of convex phase spaces, this condition is also sufficient:

6.3 Theorem
If $M \subseteq \mathbb{R}^n$ is open and simply connected (e.g., convex), then the vector field $f \in C^1(M, \mathbb{R}^n)$ is a gradient vector field if and only if ω is closed, i.e., if and only if

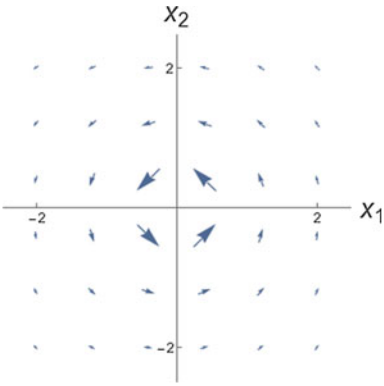
$$\frac{\partial f_i}{\partial x_k} = \frac{\partial f_k}{\partial x_i} \qquad (i, k \in \{1, \ldots, n\}). \tag{6.1.2}$$

Proof:
• The necessity of the condition follows from $f_j = \frac{\partial H}{\partial x_j}$ and the commutativity of partial derivatives of $H \in C^2(M, \mathbb{R})$.
• The converse follows from the Poincaré Lemma B.48. □

The assumption for simple connectedness of M cannot be omitted:

6.4 Example (**A Curl Free Vector Field**)
The vector field $f \in C^1(M, \mathbb{R}^2)$, $f(x) := \frac{1}{\|x\|^2} \binom{-x_2}{+x_1}$ on the phase space $M := \mathbb{R}^2 \backslash \{0\}$ has partial derivatives $\frac{\partial f_1}{\partial x_2}(x) = \frac{x_2^2 - x_1^2}{\|x\|^4} = \frac{\partial f_2}{\partial x_1}(x)$, hence satisfies (6.1.2).

The vector field f is orthogonal to the radial direction, and $\|f(x)\| = \frac{1}{\|x\|}$. Therefore, the line integral $H(x) := \int_0^1 \langle f(\gamma_x(t)), \gamma_x'(t)\rangle \, dt \quad (x \in M)$ depends only on the angle swiped out by $\gamma_x : [0, 1] \to M$. On the simply connected domain (plane with a cut)



$$\tilde{M} := \mathbb{R}^2 \setminus \big(\{0\} \times [0, \infty)\big) \subset M\,,$$

the integral is therefore the angle between the start and end points. So $f\restriction_{\tilde{M}}$ is a gradient vector field.

However, H cannot be defined continuously on M. So f itself is not a gradient vector field. ◊

The orbits of the vector field in the preceding example are circles centered at the origin. For gradient vector fields, this could not happen by Lemma 6.1.

Linear Gradient Vector Fields

Which *linear* vector fields are gradient vector fields? By Theorem 6.3, for $f(x) = Ax$, the condition $\frac{\partial f_i}{\partial x_k} = \frac{\partial f_k}{\partial x_i}$, i.e., $A_{i,k} = A_{k,i}$, has to hold for all pairs of indices (i, k): So the linear gradient vector fields are distinguished by having a self-adjoint system matrix. Then, f is the gradient of the function

$$H : \mathbb{R}^n \to \mathbb{R} \quad , \quad H(x) := \tfrac{1}{2}\langle x, Ax\rangle\,,$$

i.e., a quadratic form on the phase space $\mathbb{R}^n$.

By means of a rotation, we can achieve that the system matrix is diagonal, so, without loss of generality,

$$A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \quad \text{with} \quad \lambda_i \in \mathbb{R}\,. \tag{6.1.3}$$

Then $\exp(At) = \mathrm{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})$.

If all λ_i are nonzero, the phase space splits into the direct sum $\mathbb{R}^n = E^s \oplus E^u$ of two subspaces, a splitting that is invariant under A, and $A\restriction_{E^s} < 0$, $A\restriction_{E^u} > 0$, hence $\mathrm{Ind}(A) = \dim(E^s)$. One has

$$E^s = \Big\{x \in \mathbb{R}^n \mid \lim_{t\to\infty} \exp(At)\,x = 0\Big\}\,,\ E^u = \Big\{x \in \mathbb{R}^n \mid \lim_{t\to-\infty} \exp(At)\,x = 0\Big\},$$

so E^s and E^u are the *stable* and *unstable subspace* from Def. 5.2 respectively.

6.5 Example (**Linear Gradient Flows in the Plane**)
As the discriminant $D(A) = \mathrm{tr}(A)^2 - 4\det(A)$ of a symmetric matrix $A \in \mathrm{Mat}(2, \mathbb{R})$ is non-negative, the system matrices of gradient flows in Figure 5.3.1 on page 87 are located below the parabola $D = 0$. Conversely, each point below this parabola can be realized by a gradient flow.

After the diagonalization (6.1.3), one has $A = \mathrm{diag}(\lambda_1, \lambda_2)$, and the solutions are $x(t) = \left(e^{\lambda_1 t} x_1(0), e^{\lambda_2 t} x_2(0)\right)$, see Figure 6.1.2. ◊

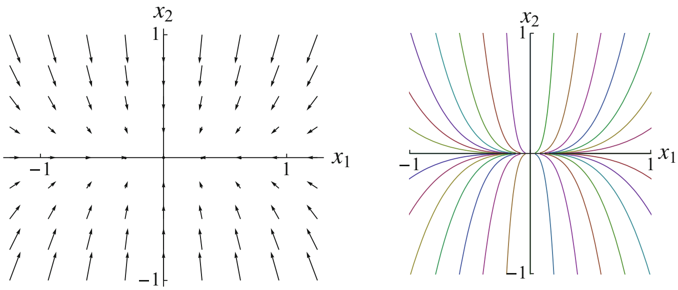


Figure 6.1.2 Left: gradient vector field $\nabla H(x)$ with $H(x) = -x_1^2 - 0.3x_2^2$. Right: its orbits

On Riemannian manifolds, the Morse functions, studied in Appendix G, give rise to gradient vector fields that are important in topological considerations.

6.1.2 *Hamiltonian Systems*

A small but significant change in the system of differential equations leads to the notion of Hamiltonian system, which is center stage in mechanics:

6.6 Definition *Let $M \subseteq \mathbb{R}^{2n}$ be open and $H \in C^2(M, \mathbb{R})$. The ODE system*

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}(p, q) \quad , \quad \dot{q}_i = \frac{\partial H}{\partial p_i}(p, q) \qquad (i = 1, \ldots, n),$$

or, in the coordinates $x \equiv (p_1, \ldots, p_n, q_1, \ldots, q_n) \equiv (p, q)$,

$$\dot{x} = X_H(x) \quad \textit{with the} \quad \textbf{Hamiltonian Vector Field} \, X_H := \mathbb{J} \nabla H$$

and $\mathbb{J} := \left(\begin{smallmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{smallmatrix}\right) \in \mathrm{Mat}(2n, \mathbb{R})$, is called **Hamiltonian Differential Equation**.

6.7 Remarks (**Hamiltonian Functions and Vector Fields**)

1. H is called *Hamilton function* or *Hamiltonian*, such as to indicate the relation with the above differential equation.
2. H is a common symbol for the Hamiltonian, going back to Lagrange.[1] In choosing this symbol, Lagrange could not have thought of Hamilton, as is obvious from the time of Hamilton's lifespan.[2]
3. In view of the connection $\nabla H = -\mathbb{J}X_H$ with the gradient vector field, Theorem 6.3 provides a criterion for a vector field $X : M \to \mathbb{R}^{2n}$ on a simply connected phase space M to be Hamiltonian. ◊

In the sequel, we will assume for the sake of simplicity that the vector field X_H is complete in the sense of Definition 3.21, so that H defines a dynamical system.[3]

6.8 Theorem *H is constant along any orbit of the Hamiltonian vector field X_H.*

Proof: Letting $\Phi : \mathbb{R} \times M \to M$ be the flow generated by $X_H = \mathbb{J}\nabla H$ and $y := \Phi(t, x)$, we calculate

$$\frac{\mathrm{d}}{\mathrm{d}t} H\big(\Phi(t,x)\big) = \mathrm{D}H_y\left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x)\right) = \mathrm{D}H_y\big(\mathbb{J}\,\nabla H(y)\big) = \langle \nabla H(y), \mathbb{J}\,\nabla H(y)\rangle\,.$$

But for $v := \nabla H(y) \in \mathbb{R}^{2n}$, one has

$$\langle v, \mathbb{J}v\rangle = \big\langle \mathbb{J}^\top v, v\big\rangle = -\langle \mathbb{J}v, v\rangle = -\langle v, \mathbb{J}v\rangle = 0\,.$$ □

Theorem 6.8 permits us to restrict the dynamical system to the level sets $H^{-1}(E)$ ($E \in \mathbb{R}$), which are often called *energy surfaces* or *energy shells*. By the implicit function theorem, these energy shells are submanifolds of M, whenever E is a regular value of H.

In the case $n = 1$, i.e., $M \subseteq \mathbb{R}^2$, this allows us to find the orbits for a given function H, albeit without the time parametrization:

- If $\nabla H(x) = 0$, the orbit is $\mathcal{O}(x) = \{x\}$.
- In contrast, if $\nabla H(x) \neq 0$, then the orbit $\mathcal{O}(x)$ is the connected component of the curve $\Sigma := \{y \in M \mid H(y) = H(x), \nabla H(y) \neq 0\}$ that contains x. The orientation is obtained by the direction given as a clockwise rotation by $\pi/2$ of the direction of the gradient, since $\mathbb{J}$ is exactly such a rotation.

---

[1] *Joseph Lagrange* (1736–1813), a French mathematician and physicist.

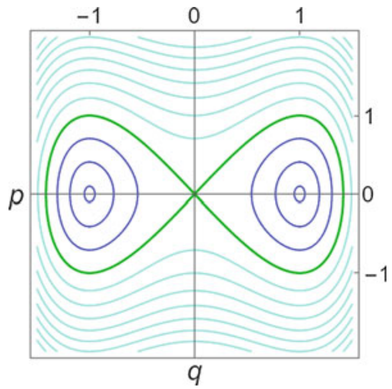
[2] *William Hamilton* (1805–1865), Irish mathematician, physicist, and astronomer.

[3] By multiplying a vector field by a suitable smooth positive function, completeness of the flow can be attained without changing the orbits. The property of being Hamiltonian will, however, not be preserved in general.

6.9 Example For the Hamiltonian

$$H : \mathbb{R}^2 \to \mathbb{R}\,,\; H(p,q) := p^2 + (q^2-1)^2\,,$$

see the figure on the right, and also Figure 6.1.1, $H^{-1}(E)$ consists of one, three or two orbits, for $E > 1$, $E = 1$ or $0 \le E < 1$, respectively. ♢



Level sets of H

Summary of Notions: For the phase space $M \subseteq \mathbb{R}^{2n} = \mathbb{R}^n_p \times \mathbb{R}^n_q$, we call

- n the *number of degrees of freedom*,
- $q \in \mathbb{R}^n_q$ the *position* and $p \in \mathbb{R}^n_p$ the *momentum*,
- $\dot q = \frac{\partial H}{\partial p}$ the *velocity* and $\ddot q$ the *acceleration*,
- $H : M \to \mathbb{R}$ the *Hamilton function* or *Hamiltonian*, or also the *total energy*.
- If the phase space is of the form $M = \mathbb{R}^n_p \times N$ with $N \subseteq \mathbb{R}^n_q$, then N is called the *configuration space*.

6.2 The Symplectic Group

6.2.1 Linear Hamiltonian Systems

For the Hamiltonian ODEs to become linear, the Hamilton function has to be of the form

$$H : \mathbb{R}^{2n} \to \mathbb{R} \quad , \quad H(x) = H(0) + \tfrac{1}{2}\,\langle x, Ax\rangle$$

with $A \in \mathrm{Mat}(2n,\mathbb{R})$ and, with no loss of generality, $A = A^\top$.

- Then it follows that $\nabla H(x) = Ax$ and $\dot x = u\,x$ with the system matrix $u := \mathbb{J}A \in \mathrm{Mat}(2n,\mathbb{R})$.
- The differential equations are invariant if we add a constant to H. This matches the fact from classical physics that we cannot measure absolute energy, but only differences in energy. Let us assume, for simplicity, that $H(0) = 0$.
- The system matrix u satisfies the identity

$$u^\top \mathbb{J} + \mathbb{J}\,u = A^\top \mathbb{J}^\top \mathbb{J} + \mathbb{J}^2 A = A - A = 0\,.$$

This leads to the following definition[4]:

[4]In the book [Art] by E. ARTIN, there is a discussion of the symplectic algebra over arbitrary fields rather than the field $\mathbb{R}$ of real numbers.

6.10 Definition (**Symplectic Algebra**) *A matrix* $u \in \mathrm{Mat}(2n, \mathbb{R})$ *and the corresponding endomorphism of* $\mathbb{R}^{2n}$ *are called* **infinitesimally symplectic** *, if*

$$u^\top \mathbb{J} + \mathbb{J}\, u = 0\,. \tag{6.2.1}$$

As the condition on u is linear, the infinitesimally symplectic endomorphisms form a subspace $\mathfrak{sp}(2n) \subset Lin(\mathbb{R}^{2n})$.

6.11 Theorem $tr(u) = 0$ *for* $u \in \mathfrak{sp}(2n)$.

Proof: $\mathrm{tr}\,(u) = -\mathrm{tr}\,(u\,\mathbb{J}^2) = -\mathrm{tr}\,(\mathbb{J}\,u\,\mathbb{J}) = -\mathrm{tr}\,(u^\top) = -\mathrm{tr}\,(u)$. □

Conclusion: (Linear) Hamiltonian systems are volume preserving.

Proof: By Theorem 4.12, $\det\big(\exp(ut)\big) = \exp\big(\mathrm{tr}(u)t\big) = 1$. □
Not only does the flow of a linear Hamiltonian system have the property of preserving volumes; it satisfies the property

$$\big(\exp(ut)\big)^\top \mathbb{J}\,\exp(ut) = \mathbb{J} \qquad (t \in \mathbb{R}), \tag{6.2.2}$$

for $\big(\exp(ut)\big)^\top = \exp(u^\top t)$, hence by (6.2.1)

$$\big(\exp(ut)\big)^\top \mathbb{J} = \mathbb{J}\,\exp(-ut) = \mathbb{J}\,\big(\exp(ut)\big)^{-1}.$$

Equation (6.2.2) says that a linear Hamiltonian flow leaves the antisymmetric bilinear form

$$\boxed{\omega_0 : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \to \mathbb{R} \quad , \quad \omega_0(u, v) := \langle u, \mathbb{J}v\rangle}$$

invariant, i.e., $\omega_0\big(\Phi_t(u), \Phi_t(v)\big) = \omega_0(u, v)$, which is similar to the fact that an orthogonal transformation leaves the canonical scalar product on $\mathbb{R}^n$ invariant. Therefore mechanical motions correspond to a particular kind of geometry, which we will investigate in more detail in the next chapter.

6.2.2 Symplectic Geometry

Symplectic geometry, which underlies mechanical motion, has certain similarities with Riemannian geometry. We want to elaborate on these similarities now.

6.12 Definition *(***Bilinear Forms***) Let E be an* $\mathbb{R}$*-vector space of finite dimension n and let* $\omega : E \times E \to \mathbb{R}$ *be a bilinear form .*

- *The* **transpose** $\omega^\top$ *of* ω *is given by* $\omega^\top(e_1, e_2) := \omega(e_2, e_1)$.
- ω *is called* **symmetric***, if* $\omega^\top = \omega$, **antisymmetric** *, if* $\omega^\top = -\omega$.
- ω *induces a linear mapping* $\omega^\flat : E \to E^*$, $\omega^\flat(e_1) \cdot e_2 := \omega(e_1, e_2)$ *into the dual space* E^* *of E.*

- *ω is called* **non-degenerate**, *if $\omega^\flat(e) = 0$ only when $e = 0$.*
- *Relative to a basis $(b_1, \ldots, b_n)$ of E, the* **representing matrix** $\mathbb{J} \in \mathrm{Mat}(n, \mathbb{R})$ **of** *ω is given by $(\mathbb{J})_{ik} := \omega(b_i, b_k)$, $(i, k = 1, \ldots, n)$.*
- *The* **rank** *of ω is the rank of a representing matrix of ω (independent of the basis chosen).*
- *If ρ is a bilinear form on F, and $f \in Lin(E, F)$, then the bilinear form*

$$f^*\rho : E \times E \to \mathbb{R} \quad , \quad f^*\rho(e_1, e_2) := \rho\big(f(e_1), f(e_2)\big)$$

is called the **pull-back of** ρ **with** f.

6.13 Theorem (Normal Forms)
Let ω be a bilinear form on an n-dimensional $\mathbb{R}$-vector space E.

1. **(Sylvester's Law of Inertia)** *If ω is symmetric with rank r, then, with respect to an appropriately chosen basis, the representing matrix of ω is of the form*

$$J = diag\big(\eta_1, \ldots, \eta_r, 0, \ldots, 0\big) = \begin{pmatrix} \eta_1 & & & & \\ & \ddots & & & \\ & & \eta_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} \in \mathrm{Mat}(n, \mathbb{R}) \,,$$

where $\eta_i \in \{-1, +1\}$.

2. **(Darboux Theorem—Linear Version)** *If ω is antisymmetric with rank r, then $r = 2m$, $m \in \mathbb{N}_0$ and with respect to an appropriately chosen basis, the representing matrix of ω is of the form*

$$\mathbb{J} = \begin{pmatrix} 0 & -1\!\!1 & 0 \\ 1\!\!1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in \mathrm{Mat}(n, \mathbb{R}) \,,$$

where $1\!\!1$ denotes the identity matrix in $\mathrm{Mat}(m, \mathbb{R})$.

Proof: These statements are frequently proved in a lecture on *linear algebra.*

1. Use the *polarization identity* $\omega(e, f) = \frac{1}{4}\big(\omega(e + f, e + f) - \omega(e - f, e - f)\big)$.

   So if $\omega \neq 0$, there exists a vector $\hat{e}_1$ with $c_1 := \omega(\hat{e}_1, \hat{e}_1) \neq 0$. Let $e_1 := \hat{e}_1/\sqrt{|c_1|}$ and $\eta_1 := \omega(e_1, e_1)$.

   Now consider the subspace spanned by e_1, call it $E_1 \subset E$ and let $E_2 := \{e \in E \mid \omega(e, e_1) = 0\}$. Then $E_1 \cap E_2 = \{0\}$ and $E_1 + E_2 = E$, because for any $z \in E$,

$$z - \eta_1 \omega(z, e_1) e_1 \in E_2 \,.$$

   Now consider the restriction of ω to E_2 and continue inductively.

2. For $\omega \neq 0$, there exist $\hat{e}_1, \hat{e}_{m+1} \in E$ with $c_1 := \omega(\hat{e}_{m+1}, \hat{e}_1) \neq 0$. Let $e_1 := \hat{e}_1/c_1$ and $e_{m+1} := \hat{e}_{m+1}$. Then we have

$$\omega(e_1, e_1) = \omega(e_{m+1}, e_{m+1}) = 0 \quad \text{and} \quad \omega(e_{m+1}, e_1) = -\omega(e_1, e_{m+1}) = 1\,.$$

Let $P_1 := \text{span}(e_1, e_{m+1}) \subset E$ and

$$E_2 := \left\{ e \in E \mid \omega(e, f) = 0 \text{ for all } f \in P_1 \right\}.$$

Then $E_2 \cap P_1 = \{0\}$ and $E_2 + P_1 = E$, because

$$z + \omega(e_1, z)e_{m+1} - \omega(e_{m+1}, z)e_1 \in E_2$$

for $z \in E$. By induction, we next treat the restriction of ω to E_2 etc. □

6.14 Definition

- *A* **symplectic form** *on a finite dimensional* $\mathbb{R}$*-vector space* E *is a non-degenerate antisymmetric bilinear form*

$$\omega : E \times E \to \mathbb{R}\,.$$

- *Then the pair* (E, ω) *is called a* **symplectic vector space** .
- *If* (E, ω) *and* (F, ρ) *are symplectic vector spaces, the linear mapping* $f : E \to F$ *is called* **symplectic** *if the* pull-back *satisfies* $f^*\rho = \omega$.

6.15 Remarks

1. On the $\mathbb{R}$-vector space $E := \mathbb{R}^n_p \times \mathbb{R}^n_q$, the standard symplectic form is ω_0, given by

$$\omega_0\big((p, q), (p', q')\big) := \sum_{j=1}^{n} \big(q_j p'_j - p_j q'_j\big) \qquad \big((p, q), (p', q') \in E\big). \quad (6.2.3)$$

On E, multiplication by the matrix $\mathbb{J} = \begin{pmatrix} 0 & -1\!\!1 \\ 1\!\!1 & 0 \end{pmatrix}$ corresponds to the automorphism $E \to E$, $(p, q) \mapsto (-q, p)$. So the mapping $\text{id} : E \to \mathbb{C}^n$, $(p, q) \mapsto p + iq$ conjugates multiplication by $\mathbb{J}$ and multiplication by the imaginary unit $i \in \mathbb{C}$. This is why $\mathbb{J}$ is also called a *complex structure* . In terms of the standard scalar product

$$\langle u, v \rangle = \sum_{j=1}^{n} u_j \overline{v_j} \qquad \big(u, v \in \mathbb{C}^n\big),$$

and with the identification of E with $\mathbb{C}^n$, the symplectic form ω_0 can be written as

$$\omega_0(u, v) = \text{Im}(\langle u, v \rangle) \qquad \big(u, v \in \mathbb{C}^n\big).$$

2. Symplectic mappings $f \in \text{Lin}(E)$ are precisely those that preserve the symplectic form ω, i.e., $f^*\omega = \omega$. By Theorem 6.13, we can find a basis of E with respect

to which the representing matrix of ω equals $\mathbb{J} = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}$. If A is the representing matrix of f, then

$$A^\top \mathbb{J} A = \mathbb{J}\,. \tag{6.2.4}$$

3. While symplectic mappings do preserve volume, in general, not every volume preserving mapping is symplectic.

   For instance, consider a four dimensional vector space E with basis $e_1, \ldots, e_4$ and symplectic bilinear form ω with matrix $\mathbb{J}$. Then the linear mapping $f : E \to E$, $(e_1, e_2, e_3, e_4) \mapsto (-e_1, -e_2, e_3, e_4)$ is volume preserving, but

$$\omega\big(f(e_1), f(e_3)\big) = -\omega(e_1, e_3)\,.$$

   In contrast, in two dimensions, if an endomorphism f of $\mathbb{R}^2$ leaves the oriented area invariant, then f is indeed symplectic. This is because all matrices $A \in \mathrm{Mat}(2, \mathbb{R})$ satisfy $(A^\top \mathbb{J} A)\mathbb{J}^{-1} = \det(A)\mathbb{1}$. This shows that (6.2.4) holds if and only if the endomorphism f given by A is area preserving. ◊

6.16 Theorem *Let (E, ω) be a symplectic vector space; then the set of symplectic endomorphisms $f : E \to E$ is a group under composition, called the* **symplectic group** *$Sp(E, \omega)$.*

6.17 Remark (**Orthogonal Group**)
For a Euclidean space (E, ω) with a positive definite symmetric bilinear form ω, we get analogously the *orthogonal group* $\mathrm{O}(E, \omega)$, which consists of compositions of rotations and reflections of E, see Example E.19. ◊

Proof of Theorem 6.16:
• If f is symplectic, then $f \in \mathrm{GL}(E)$ (see (6.2.6) below for $\det f \neq 0$), hence f^{-1} exists. Then we have $(f^{-1})^*\omega = (f^*)^{-1}\omega = (f^*)^{-1}(f^*\omega) = \omega$. So f^{-1} is symplectic as well.
• For symplectic f, g, one has $(f \circ g)^*(\omega) = g^* \circ f^*(\omega) = g^*\omega = \omega$.
• Moreover, the identity map is symplectic, so $\mathrm{Sp}(E, \omega) \neq \emptyset$. □

By Theorem 6.13, $\mathrm{Sp}(E, \omega)$ is isomorphic to the group

$$\mathrm{Sp}(2n) := \mathrm{Sp}(2n, \mathbb{R}) := \mathrm{Sp}(\mathbb{R}^{2n}, \omega_0)\,, \tag{6.2.5}$$

where ω_0, with respect to the canonical basis of $\mathbb{R}^n$, is represented by the matrix $\mathbb{J}$. (In complete analogy, by Theorem 6.13, one only needs to study the orthogonal group $\mathrm{O}(n) := \mathrm{O}(\mathbb{R}^n, \omega_0)$, whose canonical inner product has the representing matrix $\mathbb{1}$, rather than $\mathrm{O}(E, \omega)$.)

6.18 Example The time-t flow $\Phi_t(x) = \exp(ut)x$, $u = \mathbb{J}A$ of a linear Hamiltonian system with Hamilton function $H(x) = \frac{1}{2}\langle x, Ax\rangle$, $A \in \mathrm{Sym}(2n, \mathbb{R})$ is an element of the symplectic group $\mathrm{Sp}(2n)$, due to formula (6.2.2). ◊

The absolute value of the determinant of a symplectic endomorphism is 1, because from (6.2.4) one obtains

$$(\det A)^2 \;=\; (\det A)^2 \det(\mathbb{J}) \;=\; \det(A^\top \mathbb{J} A) \;=\; \det(\mathbb{J}) \;=\; 1\,. \tag{6.2.6}$$

As will be shown in Exercise B.16, one even has $\det(A) = +1$.

An immediate consequence is that the product of the $2n$ complex eigenvalues of a symplectic endomorphism $f \in \mathrm{Sp}(2n)$ equals one.

6.19 Theorem (**Eigenvalues of Symplectic Endomorphisms**)

- *If $\lambda \in \mathbb{C}$ is an eigenvalue of $f \in Sp(E,\omega)$ with (algebraic) multiplicity k, then $\overline{\lambda}$, $1/\lambda$ and $1/\overline{\lambda}$ are also eigenvalues with multiplicity k.*
- *Should 1 or -1 occur as eigenvalues, their multiplicities are even.*

Proof: $n := \frac{1}{2}\dim(E)$ denotes the number of degrees of freedom of (E,ω).

- From the fact that f is real, it follows that $\overline{\lambda}$ is an eigenvalue if λ is.
- We prove that $1/\lambda$ is an eigenvalue by studying the characteristic polynomial. To this end, we choose, by the normal form theorem 6.13, a basis of E, in which ω is represented by the matrix $\mathbb{J}$ and f by the matrix A. Then for $\lambda \in \mathbb{C}\backslash\{0\}$,

$$\begin{aligned}
\det(\lambda\mathbb{1} - A) &= \det\big(\mathbb{J}(\lambda\mathbb{1} - A)\mathbb{J}^{-1}\big) = \det\big(\lambda\mathbb{1} - \mathbb{J}A\mathbb{J}^{-1}\big)\\
&= \det\big(\lambda\mathbb{1} - (A^\top)^{-1}\big) \;\; \big(\text{since } A^\top\mathbb{J}A = \mathbb{J},\ \text{or}\, \mathbb{J}A\mathbb{J}^{-1} = (A^\top)^{-1}\big)\\
&= \det\big(\lambda\mathbb{1} - A^{-1}\big) = \det\big(A^{-1}(\lambda A - \mathbb{1})\big)\\
&= \det(A^{-1})\;\det(\lambda A - \mathbb{1}) = \det(\lambda A - \mathbb{1}) = \lambda^{2n}\det(A - \lambda^{-1}\mathbb{1})\\
&= \lambda^{2n}\det(\lambda^{-1}\mathbb{1} - A)\,.
\end{aligned}$$

- So the characteristic polynomial $p_A(\lambda) := \det(\lambda\mathbb{1} - A)$ satisfies the equation

$$p_A(\lambda) = \lambda^{2n}\, p_A(1/\lambda)\,. \tag{6.2.7}$$

If there is an eigenvalue $\lambda_0 \in \mathbb{C}$ with multiplicity k, we can split it off:

$$p_A(\lambda) = (\lambda - \lambda_0)^k\, Q(\lambda)\,, \tag{6.2.8}$$

and therefore, with (6.2.7) and (6.2.8),

$$p_A\left(\frac{1}{\lambda}\right) = \lambda^{-2n} p_A(\lambda) = \lambda^{-2n}(\lambda - \lambda_0)^k Q(\lambda) = \lambda_0^k \left(\frac{1}{\lambda_0} - \frac{1}{\lambda}\right)^k \lambda^{k-2n} Q(\lambda).$$

As Q is a polynomial of degree $2n - k$ in λ, the function $\lambda \mapsto \lambda^{k-2n}\, Q(\lambda)$ is a polynomial in $1/\lambda$. So $1/\lambda_0$ is a zero of multiplicity $l \geq k$ of $p_A(1/\lambda)$.

Exchanging the roles of λ_0 and $1/\lambda_0$ shows that indeed $l = k$.

- Now $\lambda_0 = 1/\lambda_0$ if and only if $\lambda_0 \in \{\pm 1\}$.

  As there are altogether $2n$ eigenvalues and the number of those eigenvalues that are $\neq \pm 1$ is even, the combined multiplicity of the eigenvalues 1 and -1 has to be even as well. But with $\det A = 1$, it now follows that the multiplicities of 1 and -1 are separately even. □

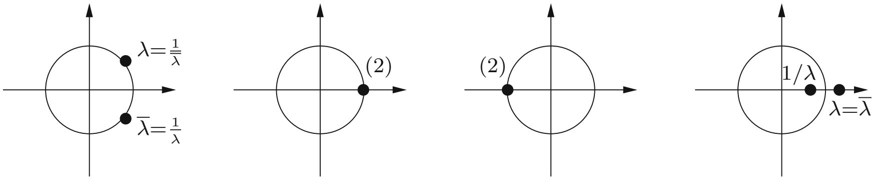


Figure 6.2.1 Complex eigenvalues of $f \in \mathrm{Sp}(2)$

When $\dim(E) = 2$, the cases that can occur are sketched in Figure 6.2.1.

As a matrix group, $\mathrm{Sp}(2n) = \{A \in \mathrm{Mat}(2n, \mathbb{R}) \mid F(A) = \mathbb{J}\}$ with

$$F : \mathrm{Mat}(2n, \mathbb{R}) \to \mathrm{Alt}(2n, \mathbb{R}) \quad , \quad A \mapsto A^\top \mathbb{J} A\,,$$

see (6.2.4). $\mathbb{J}$ is a regular value of F since, for $A \in \mathrm{Sp}(2n)$ and $B \in \mathrm{Alt}(2n, \mathbb{R})$, we have $DF_A(C) = B$ for $C := -\frac{1}{2} A\mathbb{J}B \in \mathrm{Mat}(2n, \mathbb{R})$. So the symplectic group becomes a submanifold of the $\mathbb{R}$-vector space $\mathrm{Mat}(2n, \mathbb{R})$ by Definition 2.34. This makes it even a Lie group (see Definition E.16), i.e., multiplication and inverses are smooth mappings.

Considering a *path* in $\mathrm{Sp}(E, \omega)$, i.e., a continuous mapping $c : [0, 1] \to \mathrm{Sp}(E, \omega)$, we observe that simple eigenvalues of $c(t)$ that lie on the real axis, or the unit circle, cannot leave these by changing t, see Figure 6.2.2. We will see in Theorem 7.3 that this property implies some kind of stability of linear Hamiltonian systems under small perturbations.

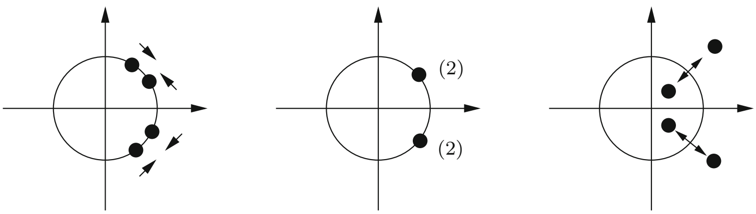


Figure 6.2.2 Complex eigenvalues of $f \in \mathrm{Sp}(4)$

6.2.3 The Symplectic Algebra

Generalizing Definition 6.10, we define:

6.20 Definition

- *A mapping* $u \in Lin(E)$ *is called* **infinitesimally symplectic** *with respect to a symplectic bilinear form* ω *if*

$$\omega\big(u(e_1), e_2\big) + \omega\big(e_1, u(e_2)\big) = 0 \qquad (e_1, e_2 \in E). \tag{6.2.9}$$

  We denote the set of these mappings as $\mathfrak{sp}(E, \omega)$.
- *The* **commutator** *of* $u, v \in Lin(E)$ *is* $[u, v] := u \circ v - v \circ u$.

Due to the linearity of (6.2.9) in u, the set $\mathfrak{sp}(E, \omega)$ is a subspace of $\mathrm{Lin}(E)$. Moreover, it even is true that

6.21 Lemma $\big(\mathfrak{sp}(E, \omega), [\cdot, \cdot]\big)$ *is a Lie algebra.*

Proof: • $\big(\mathrm{Lin}(E), [\cdot, \cdot]\big)$ is a Lie algebra, because the commutator is bilinear and alternating ($[u, u] = u \circ u - u \circ u = 0$), and the Jacobi identity holds for $u, v, w \in \mathrm{Lin}(E)$:

$$\begin{aligned}&[u, [v, w]] + [v, [w, u]] + [w, [u, v]] = \\ &\qquad u \circ v \circ w - u \circ w \circ v - v \circ w \circ u + w \circ v \circ u + v \circ w \circ u - v \circ u \circ w \\ &\qquad - w \circ u \circ v + u \circ w \circ v + w \circ u \circ v - w \circ v \circ u - u \circ v \circ w + v \circ u \circ w = 0.\end{aligned}$$

• Since for $u, v \in \mathfrak{sp}(E, \omega) \subset \mathrm{Lin}(E)$ and arbitrary $e_1, e_2 \in E$, one has

$$\begin{aligned}&\omega\big([u, v]e_1, e_2\big) + \omega\big(e_1, [u, v]e_2\big) \\ &\quad = \omega\big(u \circ v(e_1), e_2\big) - \omega\big(v \circ u(e_1), e_2\big) + \omega\big(e_1, u \circ v(e_2)\big) - \omega\big(e_1, v \circ u(e_2)\big) \\ &\quad = -\omega\big(v(e_1), u(e_2)\big) + \omega\big(u(e_1), v(e_2)\big) - \omega\big(u(e_1), v(e_2)\big) + \omega\big(v(e_1), u(e_2)\big)\end{aligned}$$

$= 0$ (using (6.2.9) for the second equation), it follows $[u, v] \in \mathfrak{sp}(E, \omega)$. □

Similar to Theorem 6.19 on the symplectic group, one obtains the following statement for $\mathfrak{sp}(E, \omega)$:

6.22 Theorem (**Symplectic Algebra**) *Let* (E, ω) *be a symplectic vector space.*

- *If* $\lambda \in \mathbb{C}$ *is an eigenvalue of* $u \in \mathfrak{sp}(E, \omega)$ *with (algebraic) multiplicity k, then* $-\lambda, \overline{\lambda}, -\overline{\lambda}$ *are also eigenvalues of multiplicity k.*
- *If 0 is an eigenvalue of* $u \in \mathfrak{sp}(E, \omega)$*, then its multiplicity is even.*

6.23 Exercise (**Symplectic Algebra**) Prove Theorem 6.22. ♢

6.24 Example $u \in \mathfrak{sp}(\mathbb{R}^2)$. In this case, the pair of eigenvalues can only be either real or purely imaginary, see Figure 6.2.3. ♢

Analogously to the mapping $\exp : \mathrm{Mat}(u, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R})$ considered in Remark 4.13, we consider the exponential mapping

$$\mathfrak{sp}(E, \omega) \to \mathrm{Sp}(E, \omega) \quad , \quad u \mapsto \exp(u)$$

from the symplectic Lie algebra

into the symplectic Lie group (see the general definition of the exponential map in (E.3.1)). Indeed, $\exp(u) \in \mathrm{Sp}(E, \omega)$, because we have, due to (6.2.9),

$$\omega\big(\exp(u)e_1\,,\,\exp(u)e_2\big) \;=\; \omega\big(e_1\,,\,\exp(-u)\exp(u)e_2\big) \;=\; \omega(e_1, e_2)\,.$$

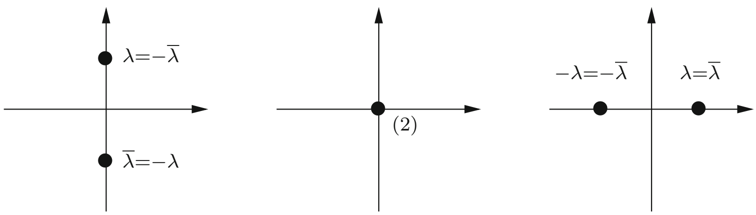


Figure 6.2.3 Eigenvalues of $u \in \mathfrak{sp}(\mathbb{R}^2)$

We see that the eigenvalues λ of u with $\mathrm{Re}(\lambda) > 0$ become eigenvalues $\exp(\lambda)$ of $\exp(u)$ with absolute value $|\exp(u)| > 1$.

6.25 Remark (**Lyapunov-Stability**)
Recalling now the fact (Example 6.18) that for a quadratic Hamiltonian $H(x) = \frac{1}{2}\langle x, Ax\rangle$, the differential equation $\dot{x} = X_H(x) = \mathbb{J}Ax$ will have the solution $\Phi_t(x) = \exp(ut)x$ with $u := \mathbb{J}A \in \mathfrak{sp}(\mathbb{R}^{2n})$, it follows from Theorem 6.22 that the fixed point 0 can only be Lyapunov-stableif all eigenvalues of u are purely imaginary. ◊

6.26 Exercise (**Symplectic Matrices**) Let $\mathbb{J} = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \in \mathrm{Mat}(2n, \mathbb{R})$.

(a) Show that $u \in \mathrm{Mat}(2n, \mathbb{R})$ is infinitesimally symplectic ($u^\top \mathbb{J} + \mathbb{J}u = 0$), if and only if $u = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ with $A, B, C, D \in \mathrm{Mat}(n, \mathbb{R})$, $D = -A^\top$ and B, C symmetric. Consequently $\dim\big(\mathfrak{sp}(2n)\big) = n(2n+1)$.

(b) Show that $a \in \mathrm{Mat}(2n, \mathbb{R})$ is symplectic, i.e., $a^\top \mathbb{J} a = \mathbb{J}$, if and only if $a = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ with $A, B, C, D \in \mathrm{Mat}(n, \mathbb{R})$, $A^\top C$ and $B^\top D$ symmetric and $A^\top D - C^\top B = \mathbb{1}$.

(c) Show by means of (b) that $\mathrm{Sp}(2, \mathbb{R}) \cong \mathrm{SL}(2, \mathbb{R})$.
Show also that $\mathrm{Sp}(2, \mathbb{R})$ is homeomorphic to the 3-dimensional solid torus $S^1 \times B$, where $B := \{x \in \mathbb{R}^2 \mid \|x\| < 1\}$ is the open disc.
Hint: Use the polar decomposition theorem from Appendix E.18:
Every matrix $M \in \mathrm{GL}(n, \mathbb{R})$ has a unique decomposition as a product $M = OP$ with $O \in \mathbf{O}(n)$ and $P \in \mathrm{Sym}(n, \mathbb{R})$ positive definite.

(d) Show that the eigenvalues of $M \in \mathrm{Sp}(2,\mathbb{R})$ are degenerate, i.e., both equal 1 or both equal -1, if and only if in the polar decomposition

$$M = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \exp\begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \text{ one has } \cos(\varphi) = \pm 1/\cosh\left(\sqrt{a^2+b^2}\right).$$

The two hypersurfaces of such *parabolic* matrices separate $\mathrm{Sp}(2,\mathbb{R})$ into four open connected components. Two of them consist of *elliptic* matrices M, i.e., such with two eigenvalues on the unit circle, other than ± 1. Two components consist of *hyperbolic* matrices, i.e., such with two real eigenvalues other than ± 1.

This is depicted in the figure at the beginning of the chapter, on page 97. ◊

6.27 Literature The normal forms of infinitesimally symplectic and of symplectic mappings under symplectic conjugacies are constructed and classified in the articles [Wil] by Williamson and [LaMe] by Laub and Meyer. ◊

6.3 Linear Hamiltonian Systems

The linear Hamiltonian vector fields $X : E \to E$ on the symplectic vector space (E,ω) correspond to elements of $\mathfrak{sp}(E,\omega)$. Without loss of generality, we assume that (E,ω) is $\mathbb{R}^{2d}$ with the symplectic structure given by $\mathbb{J} = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}$. Then the vector field can be represented in the form

$$X(x) = \mathbb{J}\nabla H(x) \quad \text{with} \quad H(x) = \tfrac{1}{2}\langle x, Ax\rangle \qquad \left(x \in \mathbb{R}^{2d}\right)$$

and $A \in \mathrm{Sym}(2d,\mathbb{R})$, hence $\nabla H(x) = Ax$.

Like any other linear differential equation with constant coefficients, the Hamiltonian initial value problem $\dot{x} = X(x)$, $x(0) = x_0$ is solved by $x(t) = \Phi_t(x_0) = \exp(Xt)x_0$.

However, hidden behind this general formula, there are a multitude of special issues, in terms of mathematics as well as of physics; we will pursue these in the present chapter.

Firstly, the motion on the energy surface $\Sigma_E := H^{-1}(E)$ is obtained (due to linearity) by scaling from the motion on the energy surfaces Σ_e, $e := \mathrm{sign}(E)$. For if $x \in \Sigma_E$ for $E \neq 0$, then $x_e := x/\sqrt{|E|} \in \Sigma_e$ and $\Phi_t(x) = \sqrt{|E|}\Phi_t(x_e)$. So it always suffices to consider merely the energies 1, 0 and -1.

6.3.1 Harmonic Oscillators

In physics, harmonic oscillators are used to describe a linear approximation to the dynamics of Hamiltonian systems for small deviations from a stable equilibrium. By

Remark 6.25, such a linear Hamiltonian vector field $X \in \mathfrak{sp}(E, \omega)$ can, in the best case, lead to Lyapunov- stability of the equilibrium 0 under the differential equation $\dot{x} = X(x)$, if all eigenvalues of X are purely imaginary.

6.28 Definition *A linear mapping $X \in Lin(E)$ is called* **semisimple** *if its minimal polynomial*[5] *does not contain quadratic factors, in other words, if the algebraic and geometric multiplicities coincide.*

6.29 Lemma *1. If $X \in \mathfrak{sp}(E, \omega)$ is semisimple and the eigenvalues of X are imaginary, then there is a* **symplectic basis** $(\hat{p}_1, \ldots, \hat{p}_d, \hat{q}_1, \ldots, \hat{q}_d)$ *of E (i.e., $\omega(\hat{p}_j, \hat{p}_k) = \omega(\hat{q}_j, \hat{q}_k) = 0$, $\omega(\hat{q}_j, \hat{p}_k) = \delta_{j,k}$) for X such that*

$$X(\hat{p}_k) = -\omega_k \hat{q}_k \quad , \quad X(\hat{q}_k) = \omega_k \hat{p}_k \qquad (k = 1, \ldots, d), \tag{6.3.1}$$

where $\pm i\,\omega_1, \ldots, \pm i\,\omega_d$ are the eigenvalues of X.

2. *Sufficient for the hypotheses in part 1 to be satisfied is that X is the Hamiltonian vector field of a positive (or negative) definite quadratic form $H : E \to \mathbb{R}$.*

Proof:

1. We certainly can write the eigenvalues of X in the form $\pm i\,\omega_k$ with $\omega_k \geq 0$, and we choose complex conjugate eigenvectors $\hat{e}_k^{\pm}$ to these eigenvalues, so $X(\hat{e}_k^{\pm}) = \pm i\,\omega_k \hat{e}_k^{\pm}$. For $\omega_k = 0$, too, there are linearly independent eigenvectors $\hat{e}_k^{\pm}$, because the multiplicity of the eigenvalue 0 is even by Theorem 6.22. Eigenspaces for eigenvalues $\omega_k \neq \omega_\ell$ are ω-orthogonal, because for $a, b \in \{-1, 1\}$ one gets (using (6.2.9))

$$0 = \omega\big(\hat{e}_k^a, (X - ib\omega_\ell \mathbb{1})\hat{e}_\ell^b\big) + \omega\big((-X + ia\omega_k \mathbb{1})\hat{e}_k^a, \hat{e}_\ell^b\big) = i(a\omega_k - b\omega_\ell)\omega(\hat{e}_k^a, \hat{e}_\ell^b).$$

If we decompose into real and imaginary part, i.e., $\hat{e}_k^{\pm} = \hat{p}_k \pm i\hat{q}_k$ with $\hat{p}_k, \hat{q}_k \in E$, then we obtain a symplectic basis by using Theorem 6.13 and rescaling. Comparing coefficients yields (6.3.1).

2. By hypothesis, $\omega(X, \cdot) = dH$. Now if H is positive definite, then this quadratic form defines the scalar product

$$\langle \cdot, \cdot \rangle_H : E \times E \to \mathbb{R} \quad , \quad \langle x, y \rangle_H = \tfrac{1}{2}\big(H(x+y) - H(x-y)\big)\,. \tag{6.3.2}$$

We choose an orthonormal basis for $\big(E, \langle \cdot, \cdot \rangle_H\big)$. With respect to this basis, we denote by $\tilde{x}$ the coordinate vector of x, and then $H(x) = \frac{1}{2}\langle x, x \rangle_H$ becomes $\frac{1}{2}\langle \tilde{x}, \tilde{x} \rangle$, and the symplectic form is represented as $\omega(x, y) = \big\langle \tilde{x}, \tilde{\mathbb{J}}\tilde{y} \big\rangle$ with $\tilde{\mathbb{J}}^\top = -\tilde{\mathbb{J}} \in \mathrm{Mat}(2d, \mathbb{R})$. So the Hamiltonian vector field X is $\tilde{x} \mapsto \tilde{\mathbb{J}}\tilde{x}$. As an anti-hermitian matrix, $\tilde{\mathbb{J}}$ is normal[6] hence semisimple; and it has purely imaginary eigenvalues.
For a negative definite quadratic form H, simply change the sign in (6.3.2). □

---

[5]This is the monic polynomial $p \in \mathbb{C}[x]$ of smallest degree for which $p(X) = 0$.

[6]$A \in \mathrm{Mat}(n, \mathbb{C})$ is called **normal**, if $A^\top A = AA^\top$.

Under the hypotheses of part 1 of Lemma 6.29, the Hamiltonian vector field X in the given basis has the Hamilton function

$$\boxed{H:\mathbb{R}^{2d}\to\mathbb{R}\quad,\quad H(p,q)=\sum_{k=1}^{d}\tfrac{\omega_k}{2}(p_k^2+q_k^2)\,.}\tag{6.3.3}$$

(6.3.3) is a normal form representation of a *harmonic oscillator* in d degrees of freedom, with the frequencies $\omega_k\in\mathbb{R}$. Its Hamiltonian equations

$$\dot p_k=-\omega_k q_k\quad,\quad \dot q_k=\omega_k p_k\qquad(k=1,\dots,d)$$

have the solutions

$$\begin{pmatrix}p_k(t)\\ q_k(t)\end{pmatrix}=\begin{pmatrix}\cos(\omega_k t)&-\sin(\omega_k t)\\ \sin(\omega_k t)&\cos(\omega_k t)\end{pmatrix}\begin{pmatrix}p_k(0)\\ q_k(0)\end{pmatrix}.\tag{6.3.4}$$

They are all bounded, and the equilibrium is Lyapunov-stable.

Now all energy surfaces $\Sigma_E=H^{-1}(E)$ are compact if H is positive or negative definite, i.e., if the frequencies ω_k all have the same sign.

6.30 Remark (**Indefinite Hamilton Function**) The case of an indefinite quadratic function (6.3.3) occurs for example as the linearization of certain problems in celestial mechanics (see Siegel and Moser [SM], §36).

In that case, it is more difficult to decide whether a small perturbation of (6.3.3) leads to a Hamiltonian dynamics whose equilibrium is still stable. ◊

Let us have a closer look at the definite case and assume without loss of generality that all frequencies are positive. The phase space functions

$$F_k:\mathbb{R}^{2d}\to\mathbb{R}^+\quad,\quad F_k(p,q):=\frac{\omega_k}{2}\left(p_k^2+q_k^2\right)\qquad(k=1,\dots,d)\tag{6.3.5}$$

are constants of motion (6.3.4), and $f=(f_1,\dots,f_d)^\top\in\mathbb{R}^d$ is a regular value in the image of

$$F:=(F_1,\dots,F_d)^\top:\mathbb{R}^{2d}\to[0,\infty)^d\tag{6.3.6}$$

if and only if the f_k are positive. In that case, $F^{-1}(f)\subset\mathbb{R}^{2d}$ is a d-dimensional submanifold of phase space, which is also invariant under the flow and diffeomorphic to a d-dimensional torus.

It is crucial to understand that qualitative properties of the dynamics depend heavily on number theoretic relations among the frequencies. We will see this in more detail in Chapter 13 on integrable systems, but will have a sneak preview already now.

6.31 Definition *The frequency vector* $\omega:=(\omega_1,\dots,\omega_d)^\top\in\mathbb{R}^d$ *is called* **rationally independent** *if no* $k\in\mathbb{Z}^d\backslash\{0\}$ *exists for which* $\langle\omega,k\rangle\equiv\sum_{i=1}^{d}\omega_i k_i=0$.

6.32 Theorem *The following are true for every positive definite Hamiltonian (6.3.3) of the harmonic oscillator and for values $E > 0$ of the energy:*

- *There are at least d periodic orbits on Σ_E.*
- *If ω is rationally independent, there are only d periodic orbits on on Σ_E.*

Proof:
• For $k = 1, \dots, d$, consider the initial data $x(0) = \big(p(0), q(0)\big) \in \Sigma_E$, given by $p_i(0) = 0$ $(i = 1, \dots, d)$, and $q_i(0) = \sqrt{\frac{2E}{\omega_k}}$ when $i = k$, and $q_i(0) = 0$ otherwise. Then the orbit through $x(0)$ is periodic with minimal period $2\pi/\omega_k$ by (6.3.4).

• If $x(0) \in \Sigma_E$ is periodic with period $T > 0$, then T must be an integer multiple of $2\pi/\omega_k$ if $F_k\big(x(0)\big) > 0$. Should this be the case for exactly one k, this corresponds to the above *normal oscillations*. Otherwise let $F_{k_i}\big(x(0)\big) > 0$ for $k_1 \neq k_2$. Then, for appropriate $\ell_{k_1}, \ell_{k_2} \in \mathbb{N}$:

$$T = 2\pi\ell_{k_1}/\omega_{k_1} \quad \text{and} \quad T = 2\pi\ell_{k_2}/\omega_{k_2} ,$$

hence $\omega_{k_1}\ell_{k_2} - \omega_{k_2}\ell_{k_1} = 0$. Thus ω is not rationally independent. □

6.33 Remarks 1. The number of only d periodic orbits on $\Sigma_E \subset \mathbb{R}^{2d}$, which occurs in the rationally independent case, is very small; it has been shown (see GINZBURG [Gi]) that large classes of nonlinear Hamiltonian differential equations do not fall below this minimum number of periodic orbits.

2. For the normal oscillations, the images of the orbit curve $t \mapsto q(t)$ in configuration space $\mathbb{R}^d$ are straight line segments along the k^{th} axis. In the case of rationally independent frequencies, the orbit curve with initial conditions $x_0 \in \mathbb{R}^{2d}$ and $f := F(x_0) \in (\mathbb{R}^+)^d$ is dense in the rectangular box

$$Q(x_0) := \left\{ q \in \mathbb{R}^d \;\middle|\; \forall k = 1, \dots, d :\ |q_k| \leq \sqrt{\tfrac{2 f_k}{\omega_k}} \right\} \tag{6.3.7}$$

(see Figure 6.3.1 left). This box is the projection of the torus $F^{-1}(f) \subset \mathbb{R}^{2d}$ on the configuration space, and the orbit is dense in $F^{-1}(f)$ by Corollary 15.8. ◊

Next, let us consider frequencies $\omega_1, \dots, \omega_d > 0$ that are all integer multiples of a single base frequency $\omega_0 > 0$. Then all orbits are closed, and $T := 2\pi/\omega_0$ is a period (not necessarily the minimal period). In this case, the projections of the trajectories in configuration space are not dense in the box (6.3.7), except in the case of a normal oscillation. For $d = 2$, the closed curves in $\mathbb{R}^2$ are called *Lissajous figures* (see Figure 6.3.1 right).

6.34 Exercises (**Lissajous Figures**) 1. How can you read off the frequency ratio $\omega_2/\omega_1 \in \mathbb{Q}$ of a harmonic oscillator from the Lissajous figure of a trajectory? (This is only possible if the oscillation is not a normal oscillation.)

2. Given a harmonic oscillator with two degrees of freedom and irrational frequency ratio ω_2/ω_1, consider for $E > 0$ a point $q \in \mathbb{R}^2$ with $H(0, q) \leq E$ for the Hamiltonian (6.3.3). Which subset of the elliptic domain $\{Q \in \mathbb{R}^2 \mid H(0, Q) \leq E\}$ can

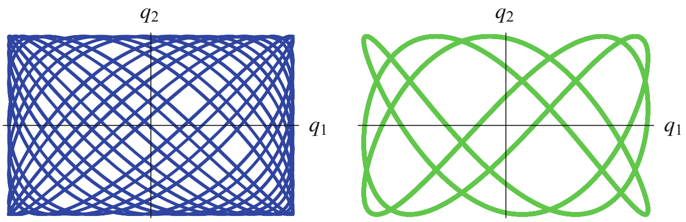


Figure 6.3.1 Lissajous figures: finite length trajectories of harmonic oscillators with two degrees of freedom, with frequency ratios $\sqrt{2}$ (left) and 5/3 (right) respectively

be reached with initial values $\{(p,q) \mid H(p,q) = E\}$?

Conclude from this example that Theorem G.15 by Hopf and Rinow about the existence of a connecting geodesic between any two given points does not have a general analog for motion within a potential. (Notice however Theorem 11.6). $\diamondsuit$

Let us begin with the simplest case $\omega_1 = \ldots = \omega_d = \omega_0$, when all frequencies are equal. Then the motion (6.3.4) is periodic with minimal period $T = 2\pi/\omega_0$, and for $E > 0$, all orbits in Σ_E are diffeomorphic to the circle S^1. We assume, without loss of generality, that $E = \frac{1}{2}\omega_0$, hence

$$\Sigma_E = S^{2d-1} = \{x \in \mathbb{R}^{2d} \mid \|x\| = 1\} \cong \{y \in \mathbb{C}^d \mid \|y\| = 1\}.$$

Passing to complex coordinates $y_k = p_k + iq_k$ $(k = 1, \ldots, d)$ permits representation of the motion (6.3.4) as

$$z(t) = e^{i\omega_0 t} z_0 \qquad \left(t \in \mathbb{R},\ z_0 \in S^{2d-1}\right). \tag{6.3.8}$$

The orbit through z_0 is therefore the intersection of the sphere S^{2d-1} with the one dimensional subspace $\mathbb{C}z_0 \subset \mathbb{C}^d$ spanned by z_0. This leads to the following statement (compare with Theorem E.36 in the appendix):

6.35 Theorem *If in the Hamiltonian (6.3.3) of the harmonic oscillator, all frequencies ω_k coincide, then the orbit space Σ_E/S^1 is the projective space $\mathbb{C}P(d-1)$.*

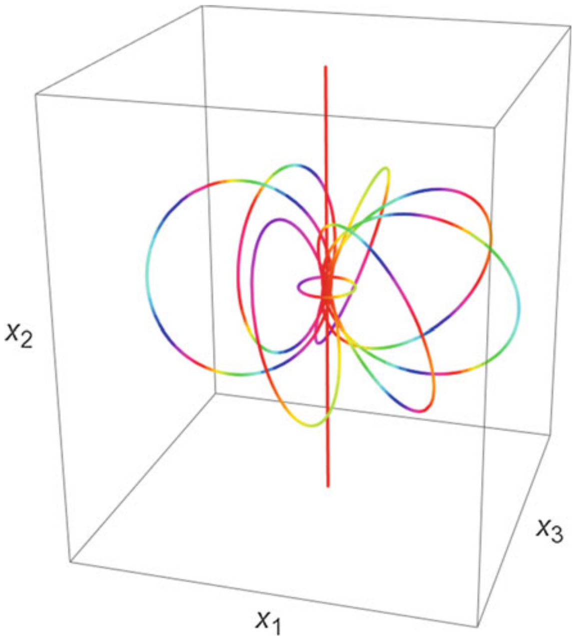
This projective space (generally defined in (E.2.1)) is a compact manifold with real dimension

$$\dim_{\mathbb{R}}\left(\mathbb{C}\mathrm{P}(d-1)\right) = 2(d-1),$$

as can be seen from the representation $\mathbb{C}\mathrm{P}(d-1) \cong S^{2d-1}/S^1$.

6.36 Remark (**Orbits of the Harmonic Oscillator for $\omega_1 = \omega_2$**) Therefore, in the case $d = 2$, the manifold of orbits on Σ_E is the 2-sphere $\mathbb{C}\mathrm{P}(1) \cong S^2$. This follows from (6.3.8) by observing that both components of $z(t) = \binom{z_1(t)}{z_2(t)}$ cannot vanish

simultaneously and that (e.g., for $z_2 \neq 0$) the ratio $\frac{z_1(t)}{z_2(t)} \in \mathbb{C}$ will be independent of time.
This ratio thus characterizes the orbit, and by means of 2-dimensional stereographic projection, the ratios $\frac{z_1}{z_2}$ and $\frac{z_2}{z_1}$ can be viewed as two coordinate systems covering the surface S^2.
By means of 3-dimensional stereographic projection, the energy surface $\Sigma_E \cong S^3$ can be identified with $\mathbb{R}^3 \cup \{\infty\}$, and thus $\mathbb{R}^3$ represents (up to one point) the orbits of the harmonic oscillator, see the figure. The surjection



$$\pi : S^3 \to S^2 \quad , \quad (z_1, z_2) \mapsto \big(2\mathrm{Re}(z_1\bar{z}_2)\,,\; 2\mathrm{Im}(z_1\bar{z}_2)\,,\; |z_2|^2 - |z_1|^2\big)$$

(with $S^3 = \{z \in \mathbb{C}^2 \mid \|z\| = 1\}$ and $S^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$) is invariant under the flow ($\pi \circ \Phi_t = \pi$) and is called *Hopf mapping*.

This mapping has a wealth of geometric applications (see also Chapter I of BATES and CUSHMAN [CB]). While for base points $o \in S^2$, the fibers $\pi^{-1}(o) \subset S^3$, which are the orbits, are of the form S^1, the bundle is not trivial, because S^3 is not diffeomorphic[7] to $S^2 \times S^1$. As can be seen in the figure (stereographically projected onto $\mathbb{R}^3$), the orbits are linked. ◊

6.37 Remark (**Linking Number**) Fortwo non-intersecting continuously differentiable closed curves $c_1, c_2 : S^1 \to \mathbb{R}^3$, the *Gauss mapping*

$$G : \mathbb{T}^2 \to S^2 \quad , \quad (t_1, t_2) \mapsto \frac{c_1(t_1) - c_2(t_2)}{\|c_1(t_1) - c_2(t_2)\|}$$

is a mapping between two compact surfaces, whose Jacobi determinant for $t := (t_1, t_2) \in \mathbb{T}^2 = S^1 \times S^1$ and $\Delta c(t) := \|c_1(t_1) - c_2(t_2)\|$ equals

$$\det\big(\mathrm{D}G(t)\big) = \det\left(G(t), \frac{c_1'(t_1)}{\Delta c(t)}, \frac{c_2'(t_2)}{\Delta c(t)}\right).$$

[7] This follows, for instance, from the observation that S^3 is simply connected, but S^1, and hence also $S^2 \times S^1$, is not, see Definition A.22.

As the surface area of the sphere S^2 equals 4π, the *linking integral*

$$LK(c_1, c_2) := \frac{1}{4\pi} \int_{\mathbb{T}^2} \det\big(DG(t)\big)\, dt_1\, dt_2 \qquad (6.3.9)$$

has an integer value. This value is called *linking number* and, intuitively speaking, describes the number of interlacings of both curves in space.[8] The quotation by Gauss printed in the box below shows that this topological notion was also inspired by a mechanical problem.

For two nonintersecting differentiable closed curves $c_1, c_2 : S^1 \to S^3$, there is always some point $n \in S^3$ that is not in the image of either curve. If we project $S^3 \backslash \{n\}$ stereographically (with respect to n) onto $\mathbb{R}^3$, we can define $LK(c_1, c_2)$ analogously. This number is independent of the choice of n because it is an integer and depends continuously on n. ◊

6.38 Exercise (**Linking Number**) For the minimal period, calculate the linking number of the two normal oscillations of a harmonic oscillator. Assume the frequencies are $\omega_1 = n_1\omega_0$, $\omega_2 = n_2\omega_0$, with relatively prime $n_1, n_2 \in \mathbb{N}$.

Using the continuity of (6.3.9), obtain the linking number of arbitrary pairs of distinct orbits in Σ_E. Compare with the figure on page 116. ◊

The Beginnings of Knot Theory

The possible geocentric positions of the orbit of a planet or asteroid make up a subset of the celestial sphere that was called *zodiacus* by Gauss.

In the case of planets, this set has the shape of a ribbon.

To elaborate on the analogous question about the newly observed asteroids, Gauss made the following distinction in his 1804 paper on astronomy:

"With respect to the location of the planetary orbit relative to the orbit of the earth, there are three cases to be distinguished. Either the latter surrounds the former, or the former the latter, or both each other (as in chain links)." …

"As can be explained by the *geometry of locations*", in the last case, the zodiacus comprises "the entire celestial sphere".

Quoted after (and translated from) M. Epple [Ep], page 65. See Figure 6.3.2

[8] For suitable orientations of $\mathbb{T}^2$ and S^2 and an arbitrary regular value $s \in S^2$ of the Gauss mapping G, the linking number equals the **mapping degree**

$$\deg(G) = \sum_{t \in G^{-1}(s)} \operatorname{sign}\big(\det(DG(t))\big).$$

.

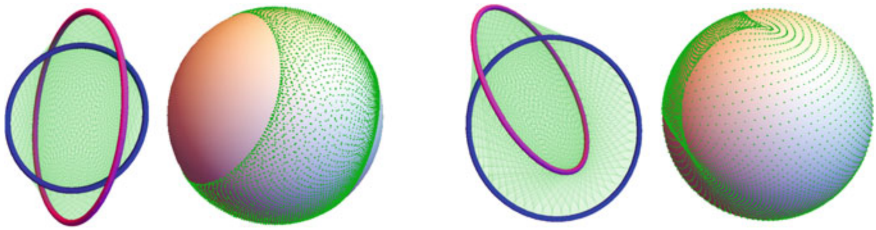


Figure 6.3.2 Zodiacus. Left: linking number 0, right: linking number 1

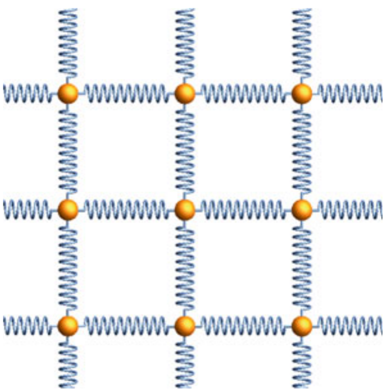
6.3.2 Harmonic Oscillations in a Lattice

We know that in principle, *linear* Hamiltonian differential equations can be solved algebraically, but in the case of many degrees of freedom, an *explicit* solution of the eigenvalue problem is often difficult or impossible.

However, in some cases, an additional symmetry of the Hamilton function helps. Oscillations of atoms in a crystal lattice about their equilibria are an example. A plausible model for the crystal is a regular lattice $\mathcal{L} \subset \mathbb{R}^d$, for example $\mathbb{Z}^3 \subset \mathbb{R}^3$, where the lattice points denote the equilibria.

But in this case, we would have to model the motion of infinitely many atoms, which is not possible with ordinary differential equations.

Instead, we use a 'finite lattice' $\mathcal{L} := (\mathbb{Z}_n)^d$ where $n \in \mathbb{N}$, and the additive group of residue classes $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z} \cong \{0, 1, \ldots, n-1\}$ (see Appendix E.1). In other words, we are using periodic boundary conditions. We start by assuming that at every lattice site $\ell \in \mathcal{L}$, there is an atom of mass $m > 0$, and that its distance from the equilibrium is $q_\ell \in \mathbb{R}^d$. Assume also that the interaction between the atoms at ℓ and $\ell' \in \mathcal{L}$ is quadratic in $\|q_\ell - q_{\ell'}\|$ and only depends on $\ell - \ell'$, i.e., is invariant under translations.



Thus on the phase space $P := \mathbb{R}_p^{dn^d} \times \mathbb{R}_q^{dn^d}$, we have the Hamilton function

$$H : P \to \mathbb{R} \quad , \quad H(p,q) = \sum_{\ell\in\mathcal{L}} \frac{\|p_\ell\|^2}{2m} + \tfrac{1}{2} \sum_{r\in\mathcal{L}} c_r \sum_{\ell\in\mathcal{L}} \|q_\ell - q_{\ell+r}\|^2 ,$$

where $c_r \geq 0$ represents the strength of the interaction. The Hamiltonian equations

$$\dot{q}_\ell = p_\ell/m \quad , \quad \dot{p}_\ell = -\sum_{r\in\mathcal{L}} c_r (2q_\ell - q_{\ell+r} - q_{\ell-r}) \qquad (\ell \in \mathcal{L}) \tag{6.3.10}$$

are solved by the complex ansatz

$$q_\ell(t) := \sum_{k\in\mathcal{L}} q_{\ell,k}(t) \quad \text{with} \tag{6.3.11}$$
$$q_{\ell,0}(t) := d_0 + d_0' t \ , q_{\ell,k}(t) := d_k \exp\bigl(2\pi i \langle k, \ell\rangle / n\bigr) \exp(i\omega_k t) \quad \bigl(k \in \mathcal{L}\backslash\{0\}\bigr)$$

(where $\langle k, \ell\rangle = \sum_{j=1}^d k_j \ell_j \pmod n$ and $d_k \in \mathbb{C}^d$).

6.39 Remark (**Translation Invariance and Fourier Transformation**)
The ansatz (6.3.11) arises from the fact that *lattice translations by the vector* $\ell \in \mathcal{L}$, i.e., the linear mappings

$$T_\ell : P \to P \quad , \quad \bigl(T_\ell(p, q)\bigr)_r = (p_{r+\ell}, q_{r+\ell}) \qquad (r \in \mathcal{L}),$$

leave the Hamilton function and thus the system matrix of the linear vector field invariant. In the complexified phase space $P_\mathbb{C} := \mathbb{C}^{2dn^d} \cong \mathbb{C}^{2d} \otimes \mathbb{C}^\mathcal{L}$, the T_ℓ are unitary mappings. These translations have the following common, mutually orthogonal eigenfunctions: $\varphi_{k,r,j} := \varphi_k\, \delta_r\, \delta_j \in P_\mathbb{C}$, where

$$\varphi_k \in \mathbb{C}^\mathcal{L} \quad , \quad \varphi_k(m) := \exp\bigl(2\pi i \langle k, m\rangle / n\bigr) \qquad (k, m \in \mathcal{L})$$

is independent of the indices $r \in \{1, 2\}$, $j \in \{1, \dots, d\}$ for the momentum and position components. Therefore the solution to the differential equation leaves the eigenspaces spanned by the $\varphi_{k,j,m}$ invariant.

The $\varphi_k : \mathcal{L} \to S^1$ $(k \in \mathcal{L})$ form the group of characters of the abelian group $\mathcal{L}$. Thus $\mathcal{L}$ is its own dual group, and we are using Fourier transformation $\mathcal{F} : \mathbb{C}^\mathcal{L} \to \mathbb{C}^\mathcal{L}$, $(\mathcal{F}\psi)(k) := \sum_{\ell\in\mathcal{L}} \psi(\ell)\varphi_k(\ell)$. ◊

The frequencies ω_k in (6.3.11) can be found by plugging the ansatz into (6.3.10): Solutions with this frequency satisfy the equations $\dot p_\ell = m\ddot q_\ell = -m\omega_k^2 q_\ell$.
On the other hand,

$$\dot p_\ell = \sum_{r\in\mathcal{L}} c_r\bigl(2q_\ell - q_{\ell+r} - q_{\ell-r}\bigr) = -\sum_{r\in\mathcal{L}} 2c_r\bigl(1 - \cos(2\pi \langle k, r\rangle / n)\bigr) q_\ell \, .$$

Altogether, this implies

$$\omega_k^2 = \frac{2}{m} \sum_{r\in\mathcal{L}} c_r\bigl(1 - \cos(2\pi \langle k, r\rangle / n)\bigr) \, , \tag{6.3.12}$$

and our ansatz was successful if $\omega_k^2 > 0$ for all $k \in \mathcal{L} \setminus \{0\}$. In the simplest case, we have a chain (i.e., $d = 1$) of n atoms in which only nearest neighbors are coupled, for instance $c_1 > 0$, $c_r = 0$ $(r \in \mathbb{Z}_n\backslash\{1\})$. One then obtains

$$\omega_k = 2\sqrt{\frac{c_1}{m}}\,|\sin(\pi k/n)|\,,$$

see Figure 6.3.3 (left).

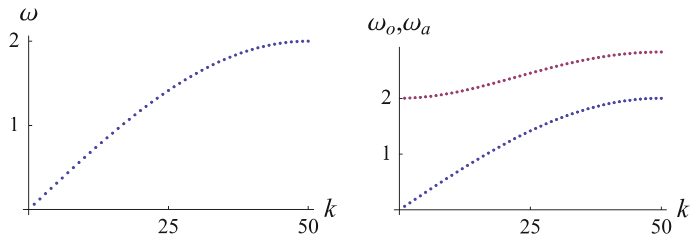


Figure 6.3.3 Dispersion relations for one dimensional lattices with $n = 100$ particles. Left: nearest neighbor coupling. Right: two atoms of different mass in the fundamental domain. Top: optical branch, bottom: acoustic branch

In the limit of a large number of particles $n \to \infty$, one obtains the relation

$$\omega(K) := 2\sqrt{\frac{c_1}{m}}\,|\sin(\pi K)|$$

in the variable $K := k/n$. It is called *dispersion relation* in physics. The derivative $\omega'(K)$ is called *group velocity* , and it can be understood as the speed with which energy is transported by waves with wave length $1/K$.

Long wave oscillations of the lattice spread here with a speed that is almost constant and proportional to $\sqrt{\frac{c_1}{m}}$, whereas for $K \to \frac{1}{2}$, the group velocity goes to 0. On page 61, at the beginning of Chapter 4, one can see the dynamics of atoms that are initially compressed like a Gaussian in the center of the image. They spread outward, so that, after some time, compression waves are moving to the left and to the right.

A more complicated picture of the dispersion relation arises if each cell of the lattice contains a number $A > 1$ of atoms.

6.40 Exercise (**Dispersion Relation**)

(a) For $A = 2$ atoms per cell of the lattice, assume that the dynamics of the chain is given by the Hamiltonian $H : \mathbb{R}^{2n}_p \times \mathbb{R}^{2n}_q \to \mathbb{R}$,

$$H(p,q) = \sum_{\ell=0}^{n-1}\left(\sum_{a=1}^{2}\frac{|p_\ell^{(a)}|^2}{2m^{(a)}} + \frac{c}{2}\left[\left(q_\ell^{(1)} - q_{\ell+1}^{(2)}\right)^2 + \left(q_\ell^{(1)} - q_\ell^{(2)}\right)^2\right]\right).$$

Here, $m^{(1)}, m^{(2)} > 0$ are the masses of the two types of atoms, and $c > 0$ is the coupling between them.

Show that the dispersion relation consists of the solutions

$$\omega_{o/a}(k) := \left(\frac{c\left(m^{(1)}+m^{(2)}\pm\sqrt{(m^{(1)}-m^{(2)})^2+m^{(1)}m^{(2)}\left(\cos\left(\frac{2\pi k}{n}\right)\left(2+\cos\left(\frac{2\pi k}{n}\right)\right)+1\right)}\right)}{m^{(1)}m^{(2)}}\right)^{\frac{1}{2}},$$

see Figure 6.3.3, right.

The solution ω_a is called *acoustic branch* in physics literature, and ω_o *optical branch*.[9]

(b) Calculate the coupling constant c_r, if the function $k \mapsto \omega_k$ from the dispersion relation (6.3.12) is given. ◊

6.41 Remarks 1. The case $k = 0$ in (6.3.11) corresponds to a motion of the entire lattice with a constant velocity.

2. In a more accurate description of physics, the lattice oscillations will be quantized and will then be called *phonons* . ◊

6.3.3 *Particles in a Constant Electromagnetic Field*

The next example leads to affine vector fields $X : \mathbb{R}^k \to \mathbb{R}^k$, i.e., those of the form

$$X(x) = Ax + b \qquad \text{with } A \in \mathrm{Mat}(k, \mathbb{R}) \text{ and } b \in \mathbb{R}^k.$$

But affine differential equations $\dot{x} = X(x)$ can just as well be solved by methods of linear algebra as the linear ones. Namely, according to the Duhamel principle (Theorem 4.20), the solution is

$$x(t) = \exp(At)\left(x_0 + \int_0^t \exp(-As)b\,\mathrm{d}s\right). \tag{6.3.13}$$

We specifically consider Hamiltonians of the form

$$H : \mathbb{R}^{2d} \to \mathbb{R} \quad , \quad H(p, q) = \tfrac{1}{2}\|p - e_0 Bq\|^2 + e_0 \langle E, q\rangle\,,$$

where $B \in \mathrm{Mat}(d, \mathbb{R})$, and $E \in \mathbb{R}^d$. In terms of physics, B is interpreted as a constant *magnetic field* and E as a constant *electric field*.[10] $e_0 \in \mathbb{R}$ is the *charge* of the particle. The equations of motion are then of the form

[9]The name is owed to the fact that, when the atoms of the two sublattices have opposite charges, an oscillation of the optical branch can generate a light wave.

[10]The linear vector field $A : \mathbb{R}^d \to \mathbb{R}^d$, $q \mapsto Bq$ is also called *vector potential* of B. The differential one form $\sum_{k=1}^d A_k(q)dq_k = \sum_{k,\ell=1}^d B_{k,\ell}\, q_\ell\, dq_k$ associated with A has exterior derivative $\sum_{k,\ell=1}^d B_{k,\ell} dq_\ell \wedge dq_k$. This two-form is called'magnetic field strength'. For more on this, including non-constant magnetic fields and the time dependent case, see Example B.21.

$$\dot{q} = p - e_0 Bq \quad , \quad \dot{p} = e_0 \left(B^\top (p - e_0 Bq) - E \right),$$

or, written as a second order differential equation, the equation of the *Lorentz force*,

$$\ddot{q} = -e_0 \big((B - B^\top)\dot{q} + E \big) . \tag{6.3.14}$$

To simplify the ensuing discussion, we assume without loss of generality that the magnetic field tensor $B \in \mathrm{Mat}(d, \mathbb{R})$ is antisymmetric; after all, only the antisymmetric part of B enters into Equation (6.3.14). (So we remark that there is no magnetic field in space dimension $d = 1$.) We also assume $e_0 = 1$.

- Firstly, if $B = 0$, then (6.3.14) has, for initial values $q(0) = q_0$, $\dot{q}(0) = v_0$, the solution

  $$q(t) = q_0 + v_0 t - \tfrac{1}{2} E t^2 .$$

  The particle is therefore accelerated in the direction of the electric field.
- In contrast, if $E = 0$, but $B \neq 0$, then
  - for $d = 2$ the antisymmetric matrix B is of the form $B = \frac{1}{2} \left(\begin{smallmatrix} 0 & b \\ -b & 0 \end{smallmatrix}\right)$, so the velocity $v(t) = \dot{q}(t)$ satisfies the differential equation $\dot{v} = b\, \mathbb{J}\, v$ with $\mathbb{J} = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and solution $v(t) = \begin{pmatrix} \cos(bt) & -\sin(bt) \\ \sin(bt) & \cos(bt) \end{pmatrix} v_0$. Therefore

    $$q(t) = q_0 + \int_0^t v(s)\, \mathrm{d}s = q_0 + \frac{1}{b} \begin{pmatrix} \sin(bt) & \cos(bt)-1 \\ 1-\cos(bt) & \sin(bt) \end{pmatrix} v_0 .$$

    This is the rotation about the center at $c := q_0 + \frac{1}{b}\, \mathbb{J}\, v_0$, with period $2\pi/|b|$ and radius $\frac{\|v_0\|}{|b|}$, also known as *Larmor radius*.
  - For even $d > 2$, we can use a theorem of Linear Algebra to transform the antisymmetric matrix B by means of an orthogonal transformation $O \in \mathrm{SO}(d)$ into the form $OBO^{-1} = \bigoplus_{k=1}^{d/2} B_k$ with $B_k = \frac{1}{2} \left(\begin{smallmatrix} 0 & b_k \\ -b_k & 0 \end{smallmatrix}\right)$; for odd d, we can likewise obtain the form $OBO^{-1} = 0 \oplus \bigoplus_{k=1}^{\frac{d-1}{2}} B_k$.

    Applying this transformation not only to the position vector $q \in \mathbb{R}^d$, but also the momentum vector $p \in \mathbb{R}^d$, leaves the form of the equations of motion invariant, and we can solve them blockwise.
  - Of all this, only the case $d = 3$ is of interest in physics.

    Then the vector $\vec{B} := \frac{1}{2} \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} \in \mathbb{R}^3$ is related to the anti-symmetric matrix $B := i(\vec{B}) := \frac{1}{2} \begin{pmatrix} 0 & -B_3 & B_2 \\ B_3 & 0 & -B_1 \\ -B_2 & B_1 & 0 \end{pmatrix}$ by

    $$\vec{B} \times \vec{V} = B\, \vec{V} \qquad \big(\vec{V} \in \mathbb{R}^3\big), \tag{6.3.15}$$

so we recognize the usual vectorial form of the magnetic field. By a rotation with a matrix $O \in \mathrm{SO}(3)$, the direction of $\vec{B}$ becomes the first coordinate direction: $OBO^{-1} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & b \\ 0 & -b & 0 \end{pmatrix}$. We will see that we have free motion in the first coordinate direction, and the already described circular motion in the $(2-3)$-plane. Altogether, the orbits of the particles $t \mapsto q(t)$ are therefore helices in $\mathbb{R}^3$.

- In dimension $d = 2$, we now consider the case of nonvanishing electric and magnetic fields.

So we write (6.3.14) as a differential equation for the velocity $w := \dot{q}$, namely $\dot{w} = b\,\mathbb{J}\,w - E$. The solution to the corresponding homogeneous equation $\dot{v} = b\,\mathbb{J}\,v$ has already been determined, namely

$$v(t) = O(t)v_0 \quad \text{with} \quad O(t) := \begin{pmatrix} \cos(bt) & -\sin(bt) \\ \sin(bt) & \cos(bt) \end{pmatrix}.$$

By (6.3.13), we have therefore

$$w(t) = O(t)\left(v_0 + \int_0^t O(-s)E\,\mathrm{d}s\right) = O(t)v_0 + \frac{1}{b}\mathbb{J}\left(\mathbb{1} - O(t)\right)E\,.$$

Integrating again yields

$$q(t) = q_0 + \int_0^t w(s)\,\mathrm{d}s = q_0 + \frac{1}{b}\mathbb{J}\left(\mathbb{1} - O(t)\right)v_0 + \frac{1}{b}\mathbb{J}\,Et + \frac{\mathbb{1} - O(t)}{b^2}E\,.$$

This is a superposition of circular motion with straight motion in direction $\mathbb{J}E$, which is *orthogonal* to the electric field, see Figure 6.3.4.

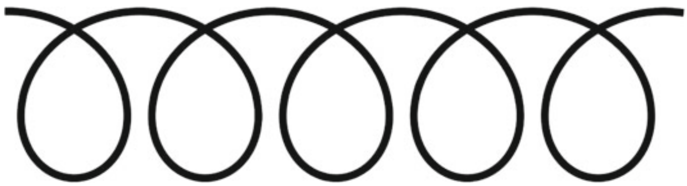


Figure 6.3.4 Motion in the plane, with constant electromagnetic fields, and vertical direction of the electric field

6.42 Exercise (**Upper Half Plane and Möbius Transformations**)
By Exercise 6.26,

$$\mathrm{Sp}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R}) = \{M \in \mathrm{Mat}(2,\mathbb{R}) \mid \det M = 1\}\,.$$

The *upper half plane* is $\mathbb{H} := \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$, and

$$\overline{\mathbb{H}} := \{z \in \mathbb{C} \mid \operatorname{Im} z \geq 0\} \cup \{\infty\}\,.$$

(a) Show that for each $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{R})$, the *Möbius transformation*

$$\widehat{M} : \mathbb{H} \to \mathbb{H} \quad , \quad \widehat{M} z := \frac{az + b}{cz + d}$$

is well-defined. (See also Example 16.17.)

(b) Show that the mapping $\mathrm{SL}(2, \mathbb{R}) \times \mathbb{H} \to \mathbb{H}$ given by $(M, z) \mapsto \widehat{M} z$ is a group operation.

(c) Möbius transformations can be extended continuously to $\overline{\mathbb{H}}$. A matrix $M \in \mathrm{SL}(2, \mathbb{R})$ with $|\operatorname{tr} M| < 2$ is called *elliptic*, one with $|\operatorname{tr} M| = 2$ *parabolic*, and one with $|\operatorname{tr} M| > 2$ *hyperbolic*.
Show: Möbius transformations corresponding to an elliptic matrix have exactly one fixed point in $\mathbb{H}$. Möbius transformations for parabolic matrices other than $\pm \mathbb{1}$ have exactly one fixed point, for hyperbolic matrices exactly two fixed points, and these fixed points lie in $\partial \mathbb{H} := \overline{\mathbb{H}} \setminus \mathbb{H}$.

(d) On $\mathbb{H}$, one defines a metric

$$g(z) := (\operatorname{Im} z)^{-2}\, g_{\mathrm{Euclid}}(z) \qquad (z \in \mathbb{H}),$$

where g_{Euclid} is the Euclidean metric on $\mathbb{H} \subset \mathbb{C}$.
Show that the Möbius transformations for $\left(\begin{smallmatrix} a & b \\ 0 & a^{-1} \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & c \\ -c^{-1} & 0 \end{smallmatrix}\right)$ with $a, b, c \in \mathbb{R}$ and $a, c \neq 0$ are isometries of g. Conclude that all Möbius transformations are isometries of g.

(e) $\mathfrak{sp}(\mathbb{R}^2)$ denotes the group of infinitesimally symplectic 2×2 matrices. Show $\mathfrak{sp}(\mathbb{R}^2) = \mathfrak{sl}(\mathbb{R}^2) := \{m \in \mathrm{Mat}(2, \mathbb{R}) \mid \operatorname{tr}(m) = 0\}$.

(f) An infinitesimally symplectic matrix $m \in \mathfrak{sp}(\mathbb{R}^2)$ generates the one-parameter group $\{\exp(tm) \mid t \in \mathbb{R}\} \subseteq \mathrm{Sp}(2, \mathbb{R})$. Which one-parameter groups are generated by

$$m \in \{\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\}\,?$$

Classify them as elliptic, parabolic, and hyperbolic respectively. Which are the fixed points of the corresponding Möbius transformations on $\overline{\mathbb{H}}$? What do the orbits $\{\widehat{\exp(tm)}\, v \mid t \in \mathbb{R}\}$ through $v \in \mathbb{H}$ look like? ◊

6.4 Subspaces of Symplectic Vector Spaces

In a vector space E, the group of structure preserving mappings is the General Linear Group $\mathrm{GL}(E)$. Two subspaces of E can be mapped into each other by a mapping in $\mathrm{GL}(E)$ if and only if their dimensions coincide.

In a symplectic vector space (E, ω), the group of structure preserving mappings is the symplectic group $\mathrm{Sp}(E, \omega)$, which is a strict subgroup of $\mathrm{GL}(E)$. For $\dim(E) \geq 4$ there are more invariants of subspaces under $\mathrm{Sp}(E, \omega)$ than merely the dimension, namely the rank of the restriction of the bilinear form ω to the subspace.

Of particular importance are those subspaces that have either rank 0 or maximal rank:

6.43 Definition *Let (E, ω) be a symplectic vector space and $F \subseteq E$ a subspace.*

- *The ω-**orthogonal complement** of F is defined to be the subspace*

$$F^\perp := \left\{ e \in E \mid \ \forall f \in F : \ \omega(e, f) = 0 \right\}.$$

- *F is called **isotropic** if $F^\perp \supseteq F$, and **Lagrangian** if $F^\perp = F$.*
- *F is called **symplectic**, if $F^\perp \cap F = \{0\}$.*

Symplectic subspaces (F, ω_F) of E with imbedding $\imath : F \to E$ and pulled-back two-form $\omega_F := \imath^*(\omega)$ are therefore symplectic vector spaces in their own right, and a subspace cannot at the same time be isotropic and symplectic (except when $\dim(E) = 0$).

6.44 Examples We consider the standard case for a $2n$-dimensional symplectic vector space, namely $(E, \omega_0) = \left(\mathbb{R}^n_p \times \mathbb{R}^n_q, \left\langle \cdot, \left(\begin{smallmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{smallmatrix}\right) \cdot \right\rangle\right)$.

1. For $0 \leq k, \ell \leq n$ and $F := \left(\mathbb{R}^k_p \times \{0\}^{n-k}\right) \times \left(\{0\}^{n-\ell} \times \mathbb{R}^\ell_q\right)$ the space $F^\perp$ is

$$F^\perp = \left(\mathbb{R}^{n-\ell}_p \times \{0\}^\ell_q\right) \times \left(\{0\}^k \times \mathbb{R}^{n-k}_q\right).$$

   A subspace F of this form is isotropic if and only if $k + \ell \leq n$, and is Lagrangian if and only if $k + \ell = n$. In particular, the subspaces $\mathbb{R}^n_p \times \{0\}_q$ and $\{0\} \times \mathbb{R}^n_q$ of E are Lagrangian.
2. The one-dimensional subspaces are isotropic because ω_0 is antisymmetric. If $\dim E = 2$, then it is exactly the one dimensional subspaces that are Lagrangian, because ω_0 ist not degenerate.
3. For a matrix $A \in \mathrm{Mat}(n, \mathbb{R})$, let $F_A := \{(q, Aq) \mid q \in \mathbb{R}^n\}$, i.e., the graph of the linear mapping represented by A. This n-dimensional subspace is Lagrangian if and only if $0 = \omega_0\big((q, Aq), (q', Aq')\big) = \langle q, Aq' \rangle - \langle Aq, q' \rangle$ for all q, q'; in other words, if A is symmetric (see Theorem 6.46.2).
4. For the subspaces $F := \left(\mathbb{R}^k_p \times \{0\}^{n-k}\right) \times \left(\mathbb{R}^k_q \times \{0\}^{n-k}\right)$, one has

$$F^\perp = \left(\{0\}^k \times \mathbb{R}^{n-k}_p\right) \times \left(\{0\}^k \times \mathbb{R}^{n-k}_q\right),$$

   so they are symplectic. $\Diamond$

6.45 Exercise (**Symplectic Mappings and Subspaces**) Let (E, ω) be a symplectic vector space.

(a) Show that for any two vectors $v, w \in E \setminus \{0\}$, there is a symplectic mapping $f \in \mathrm{Sp}(E, \omega)$ for which $f(v) = w$.
(b) Show by counterexample that, for $\dim(E) > 2$, an arbitrary 2-dimensional subspace F of (E, ω) cannot be mapped by symplectic mapping $f \in \mathrm{Sp}(E, \omega)$ on an arbitrary 2-dimensional subspace F'.
(c) Show that it is possible to map any symplectic subspace F onto any other symplectic subspace F' of the same dimension, $\dim(F') = \dim(F)$, by some symplectic mapping $f \in \mathrm{Sp}(E, \omega)$. ♢

6.46 Theorem *Let (E, ω) be a symplectic vector space and $F \subseteq E$ a subspace. Then*

1. $\dim(F) + \dim(F^\perp) = \dim(E)$.
2. $F \subseteq E$ is Lagrangian if and only if F is isotropic and $\dim(F) = \frac{1}{2}\dim(E)$.

Proof:

1. We reduce the dimension formula to the one for orthogonal subspaces. According to the normal form theorem 6.13.2, we may assume that $E = \mathbb{R}^{2n} = \mathbb{R}^n_p \times \mathbb{R}^n_q$ and ω is of the form (6.2.3). Then, with respect to the Euclidean inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^{2n}$, one has the formula $\omega_0(X, Y) = \langle X, \mathbb{J}Y\rangle$ with $\mathbb{J} = \left(\begin{smallmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{smallmatrix}\right)$.

   For $n = 1$, $\mathbb{J}$ is a rotation of the plane E by $\pi/2$, for $n > 1$ it is a rotation by $\pi/2$ in all (p_i, q_i)-planes. Then
   $$F^\perp = \{X \in E \mid \forall\, Y \in F : \langle X, \mathbb{J}Y\rangle = 0\}$$
   is the subspace that is orthogonal to $\mathbb{J}F$ with respect to $\langle\cdot,\cdot\rangle$; this proves the first claim.
2. If $\dim F = \frac{1}{2}\dim E$, and hence by the first part also $\dim F^\perp = \frac{1}{2}\dim E$, then it follows from the assumed isotropy ($F \subseteq F^\perp$) of F that $F = F^\perp$ already. Conversely, if F is Lagrangian, hence by definition isotropic, it follows from part 1 that $\dim F = \frac{1}{2}\dim E$. □

6.47 Exercise (**Dimension Formula**) In the proof of Theorem 6.46, part 1, the antisymmetry of ω was used. Construct a proof that does not use antisymmetry and relies only on the non-degenerate property of ω, so it can also be used for orthogonal spaces with respect to the scalar product (or a Lorentz metric). ♢

Graphs of symplectic mappings can be viewed as Lagrangian subspaces:

6.48 Theorem *Let* (E_1, ω_1) *and* (E_2, ω_2) *be symplectic vector spaces and* $\pi_i : E_1 \oplus E_2 \to E_i$, $i = 1, 2$, *the projections. Then*

$$\omega_1 \ominus \omega_2 := \pi_1^* \omega_1 - \pi_2^* \omega_2 .$$

is a symplectic form on the vector space $E_1 \oplus E_2$.

Proof: We need to show that the antisymmetric bilinear form $\omega := \omega_1 \ominus \omega_2$ on $E := E_1 \oplus E_2$ is not degenerate. A vector $e = (e_1, e_2) \in E \setminus \{0\}$ must satisfy $e_1 \neq 0$ or $e_2 \neq 0$. In the former case, there is a vector $f_1 \in E_1$ with $\omega_1(e_1, f_1) \neq 0$, because (E_1, ω_1) is a symplectic vector space. Thus $f := (f_1, 0)$ satisfies $\omega(e, f) = \omega_1(e_1, f_1) \neq 0$. The case $e_2 \neq 0$ is analogous. □

6.49 Theorem *An isomorphism* $A : E_1 \to E_2$ *is symplectic if and only if its graph*

$$\Gamma_A = \{(e_1, Ae_1) \mid e_1 \in E_1\} \subset E_1 \oplus E_2$$

is a Lagrangian subspace of the symplectic vector space $\big(E_1 \oplus E_2, \omega_1 \ominus \omega_2\big)$.

Proof: Since the linear mapping A is an isomorphism, $\dim(E_1) = \dim(E_2)$. The graph Γ_A thus being a linear subspace of dimension $\dim(\Gamma_A) = \frac{1}{2} \dim(E_1 \oplus E_2)$, it is Lagrangian if and only if it is isotropic, i.e.,

$$\omega_1 \ominus \omega_2\big((e_1, Ae_1), (e_1', Ae_1')\big) = 0 \qquad (e_1, e_1' \in E_1),$$

or equivalently

$$\omega_1(e_1, e_1') - \omega_2(Ae_1, Ae_1') = 0 \qquad (e_1, e_1' \in E_1).$$

This is the case exactly if A is symplectic. □

6.5 *The Maslov Index

Lagrangian subspaces frequently occur as tangential spaces of flow invariant tori in integrable Hamiltonian systems. An example for such invariant tori is the pre-image $F^{-1}(f)$ for F as in (6.3.6)). In such cases, we will consider the dependence of the Lagrangian subspace on the point of the torus. In preparation, let us investigate the set of *all* Lagrangian subspaces (without orientation).

6.50 Definition

- *For an* $\mathbb{R}$*-vector space*[11] v *with* $m := \dim(v) < \infty$*, we call the sets*

$$Gr(v,n) := \{u \subseteq v \mid u \text{ is an } n\text{-dimensional subspace}\} \quad \big(n \in \{0,\ldots,m\}\big)$$

  the **Grassmann manifolds** *of* v*, and let* $Gr(m,n) := Gr(\mathbb{R}^m, n)$.
- *Specifically, we call* $\mathbb{R}P(m-1) := Gr(m,1)$ *the* **projective space**.
- *For a symplectic vector space* (E,ω)*, the set*

$$\Lambda(E,\omega) := \{u \subseteq E \mid u \text{is Lagrangian subspace}\}$$

  is called **Lagrange-Grassmann manifold** *of* E*, and* $\Lambda(m) := \Lambda\left(\mathbb{R}^{2m},\omega_0\right)$.

Since all m-dimensional $\mathbb{R}$-vector spaces are isomorphic, it suffices to study the $\mathrm{Gr}(m,n)$; analogously, it suffices to study $\Lambda(m)$, since all $2m$-dimensional symplectic vector spaces are isomorphic (see Exercise 6.45 (c)).

6.51 Theorem *In a natural way,*

- $Gr(m,n)$ *is a compact manifold of dimension* $n(m-n)$*, with*

$$Gr(m,m-n) \cong Gr(m,n)\,,$$

- *and* $\Lambda(n) \subseteq Gr(2n,n)$ *is a compact submanifold of dimension* $\frac{1}{2}n(n+1)$.

6.52 Examples (**Grassmann Manifolds**)

1. According to Example 6.44.2, all one-dimensional subspaces of a two dimen - sional symplectic vector space are Lagrangian, hence $\Lambda(1) = \mathrm{Gr}(2,1) = \mathbb{R}\mathrm{P}(1)$.

   The one dimensional subspaces (without orientation) of $\mathbb{R}^2$ are parametrized by their angle $\varphi \in [0,\pi]$ with respect to the first coordinate axis, where $\varphi = 0$ and $\varphi = \pi$ represent this coordinate axis itself. This way, the projective space $\mathbb{R}\mathrm{P}(1)$ is naturally homeomorphic to S^1.
2. For $\left(\mathbb{R}^4,\omega_0\right)$, the four-dimensional manifold $\mathrm{Gr}(4,2)$ is the disjoint union of the three-dimensional submanifold $\Lambda(2)$ and the open and dense set of two-dimensional symplectic subspaces in $\mathrm{Gr}(4,2)$. ◊

Proof of Theorem 6.51:

- To begin with, we may view $\mathrm{Gr}(m,n)$ as a subset of the vector space $\mathrm{Lin}(\mathbb{R}^m)$, since there is a one-to-one correspondence between the n-dimensional subspaces $u \subset \mathbb{R}^m$ and the orthogonal projections $P_u \in \mathrm{Lin}(\mathbb{R}^m)$ onto them.
  This makes $\mathrm{Gr}(m,n)$ into a metric space with the distance

$$d(u,v) := \|P_u - P_v\|\,.$$

[11] Analogous definitions apply to $\mathbb{C}$-vector spaces.

- Now $\mathrm{Gr}(m,n) \subset \mathrm{Lin}(\mathbb{R}^m)$ becomes a manifold if, on the neighborhood

$$U := \{v \in \mathrm{Gr}(m,n) \mid d(u,v) < 1\}$$

of $u \in \mathrm{Gr}(m,n)$, we define the coordinate map[12] Φ implicitly by

$$\Phi : U \to \mathrm{Lin}(u, u_s) \quad , \quad v = \mathrm{Graph}\big(\Phi(v)\big)\,, \tag{6.5.1}$$

where u_s is the subspace that is orthogonal to u with respect to the Euclidean scalar product:

$$u_s := \{x \in \mathbb{R}^m \mid \forall y \in u : \langle x, y\rangle = 0\}\,.$$

In doing so, the condition $\|P_v - P_u\| < 1$ guarantees that only the zero vector in v is orthogonal to the subspace u, hence v is the graph of of a uniquely determined linear mapping from $\mathrm{Lin}(u, u_s)$. The space u_s, which is orthogonal to the n-dimensional subspace u, has dimension $m - n$, hence $\dim\big(\mathrm{Gr}(m,n)\big) = \dim\big(\mathrm{Lin}(u, u_s)\big) = n(m-n)$.
Being the set of orthogonal projections of rank n, the Grassmannian $\mathrm{Gr}(m,n)$ is bounded (with bound 1 in the operator norm) and closed, hence compact.
- The mapping $\mathrm{Gr}(m,n) \to \mathrm{Gr}(m, m-n)$, $P_u \mapsto \mathbb{1} - P_u$ is a diffeomorphism.
- The Lagrangian subspaces u of $\mathbb{R}^{2n}$ are distinguished among all n-dimensional subspaces by the condition $u^\perp = u$, and thus they form a closed (and thus compact) subset of $\mathrm{Gr}(2n,n)$. In Example 6.44.3, we had already observed that the graph $\{(q, Aq) \mid q \in \mathbb{R}^n\} \subset E = \mathbb{R}^n \times \mathbb{R}^n$ of a matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ is a Lagrangian subspace if and only if A is symmetric. Analogously, for $A \in \mathrm{Lin}(u, u_s)$, the graph $\mathrm{graph}(A)$ is a Lagrangian subspace exactly if

$$\omega\big(x + A(x), y + A(y)\big) = 0 \qquad (x, y \in u). \tag{6.5.2}$$

Since u and $u_s (= \mathbb{J}u)$ are Lagrangian subspaces,

$$\omega(x, y) = 0 \quad \text{and} \quad \omega\big(A(x), A(y)\big) = 0\,.$$

Therefore, under the hypothesis (6.5.2), one has $\omega\big(A(x), y\big) = \omega\big(A(y), x\big)$, which means that the bilinear form

$$u \times u \to \mathbb{R} \quad , \quad (x, y) \mapsto \omega\big(A(x), y\big)$$

is symmetric. As $\dim(u) = n$, we obtain an $n(n+1)/2$-dimensional space of such $A \in \mathrm{Lin}(u, u^\perp)$, and therefore the claimed formula about the dimension. □

[12] Strictly speaking, Φ becomes a coordinate map only when we identify $\mathrm{Lin}(u, u_s)$ with $\mathrm{Mat}\big(n \times (m-n), \mathbb{R}\big) \cong \mathbb{R}^{n(m-n)}$ by choosing a basis. Moreover, we use in (6.5.1) that $\mathbb{R}^n = u \oplus u_s$.

6.53 Exercise (SO(3) $\cong$ $\mathbb{R}$P(3))
Find a diffeomorphism (i.e., an invertibly smooth mapping) between the rotation group SO(3) (see Example E.28) and the projective space $\mathbb{R}$P(3). ◊

Following work by V.P. Maslov, V.I. ARNOL'D defined smooth mappings

$$\mathrm{MA}_m : \Lambda(m) \to S^1 \qquad (m \in \mathbb{N}) \tag{6.5.3}$$

in [Ar4]. In the simplest case $\Lambda(1) \cong S^1$, this is the identity map.[13] The definition for general m will follow in Lemma 6.57. These mappings play an important role in the study of integrable quantum mechanical systems, for instance the approximate calculation of eigenvalues, which leads to the *Maslov index* [14] as a correction term to the Bohr-Sommerfeld quantization. Continuous mappings $c : S^1 \to \Lambda(m)$ can be concatenated with MA_m to a continuous mapping

$$\mathrm{MA}_m \circ c : S^1 \to S^1 . \tag{6.5.4}$$

6.54 Definition

- *The* **mapping degree** *of a mapping* $f \in C^1(S^1, S^1)$ *is*[15]

$$\deg(f) := \int_{S^1} \frac{\mathrm{d}}{\mathrm{d}z} \log\big(f(z)\big) \frac{\mathrm{d}z}{2\pi i} .$$

- *The* **Maslov index** *of a continuously differentiable mapping* $c : S^1 \to \Lambda(m)$ *is defined as* $\deg(MA_m \circ c)$.

In other words, the mapping degree of $f : S^1 \to S^1$ measures the number of roundtrips made by $f(z)$ per one roundtrip of z. It is invariant under orientation preserving changes of parameter; under orientation reversing changes of parameter, however, it changes sign.

6.55 Examples (Mapping degree and Maslov Index)

1. For $m \in \mathbb{Z}$, the mapping $f_m : S^1 \to S^1$, $f(z) = z^m$ has mapping degree $\deg(f_m) = \int_{S^1} \frac{m}{z} \frac{\mathrm{d}z}{2\pi i} = m$.

   If $f : S^1 \to S^1$ is of the form $f(z) = e^{i\varphi(z)} f_m(z)$ with $\varphi \in C^1(S^1, \mathbb{R})$, then it is still true that $\deg(f) = m$, due to the fundamental theorem of calculus.

[13] If both the points on S^1 and the subspaces $\mathbb{R} \subset \mathbb{R}^2$ are parametrized by an angle φ as in Example 6.52, this identity map takes the form $\mathrm{MA}_1 : \Lambda(1) \to S^1$, $\varphi \mapsto 2\varphi$!

[14] See also BOTT [Bo1].

[15] We consider the circle as $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ and extend the natural logarithm in such a way that the mapping $S^1 \to i\mathbb{R}$, $z \mapsto \frac{\mathrm{d}}{\mathrm{d}z} \log\big(f(z)\big)$ is continuous.

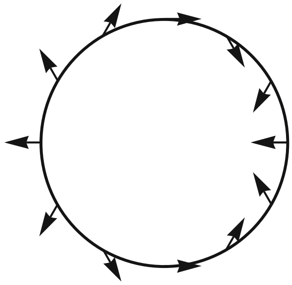
2. If we parametrize the level set $H^{-1}(E)$ of the harmonic oscillator

$$H : \mathbb{R}_p \times \mathbb{R}_q \to \mathbb{R} \quad , \quad H(p,q) = \tfrac{1}{2}(p^2+q^2)$$

for $E > 0$ as the path $\tilde{c} : S^1 \to H^{-1}(E)$, $\tilde{c}(z) := \sqrt{2E}z$ (where we have identified $\mathbb{R}_p \times \mathbb{R}_q$ with $\mathbb{C}$), then the tangent space of $H^{-1}(E)$ at $\tilde{c}(z)$ is spanned by the vector $i\tilde{c}(z)$.
We obtain a mapping

$$c : S^1 \to \Lambda(1) \quad , \quad c(z) = \operatorname{span}\bigl(i\tilde{c}(z)\bigr)$$

with Maslov index $\deg(\mathrm{MA}_1 \circ c) = 2$.
The non-oriented tangent spaces at $(-p,-q)$ are the same as at $(p,q) \in H^{-1}(E)$; see also the adjacent figure. ◊



On the Maslov index of the 1D harmonic oscillator.

Now for the definition of the mapping MA_m in (6.5.3): This leads us to the following representation of the Lagrange-Grassmann manifolds.

6.56 Theorem (**$\Lambda(m)$ as a Homogenous Space**) *For all $m \in \mathbb{N}$, the map*

$$U(m)/O(m) \to \Lambda(m) \quad , \quad U\,O(m) \mapsto \left\{ \begin{pmatrix} Re(U\,O)\,x \\ Im(U\,O)\,x \end{pmatrix} \;\middle|\; O \in O(m),\ x \in \mathbb{R}^m \right\} \tag{6.5.5}$$

is a homeomorphism between the homogenous space[16] *$U(m)/O(m)$ and the Lagrange-Grassmann manifold $\Lambda(m)$.*

Proof:
• For $V \in \mathrm{U}(m)$, the set $\left\{ \begin{pmatrix} \mathrm{Re}(V)\,x \\ \mathrm{Im}(V)\,x \end{pmatrix} \;\middle|\; x \in \mathbb{R}^m \right\}$ is a Lagrangian subspace: Any m-dimensional subspace $L \subset \mathbb{R}^m \times \mathbb{R}^m$ can be represented as the image of an injective linear map

$$\mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m \quad , \quad x \mapsto \begin{pmatrix} Ax \\ Bx \end{pmatrix}$$

with $A, B \in \mathrm{Mat}(m, \mathbb{R})$ and $\operatorname{rank}\begin{pmatrix} A \\ B \end{pmatrix} = m$. Then L is a Lagrangian subspace if and only if

$$\left\langle \begin{pmatrix} Ay \\ By \end{pmatrix}, \mathbb{J}\begin{pmatrix} Ax \\ Bx \end{pmatrix} \right\rangle = 0 \qquad \bigl(x, y \in \mathbb{R}^m\bigr),$$

i.e., if $A^\top B = B^\top A$.

[16]**Definition:** A set M is called a *homogenous space* , if some group G operates transitively on M. —In the present example, the Lie group $\mathrm{U}(m)$ operates continuously on the set $\mathrm{U}(m)/\mathrm{O}(m)$ of equivalence classes of unitary matrices if we use the quotient topology coming from $\mathrm{U}(m)$ (see page 484).

The normalization requirement that the images of the basis vectors $e_1, \dots, e_m \in \mathbb{R}^m$ should be an orthonormal basis of L is tantamount to $(A^\top \, B^\top)\binom{A}{B} = A^\top A + B^\top B = \mathbb{1}$.

Both requirements are satisfied together if and only if $V := A + iB$ is unitary, because

$$V^* V = (A^\top - iB^\top)(A + iB) = (A^\top A + B^\top B) + i(A^\top B - B^\top A) = \mathbb{1}\,.$$

• A given Lagrangian subspace still corresponds to many unitary matrices V, because we could have taken any orthonormal basis of $\mathbb{R}^m$ instead of the basis vectors $e_1, \dots, e_m \in \mathbb{R}^m$. Choosing such an orthonormal basis however corresponds to a change of basis by means of an orthogonal matrix from $\mathrm{O}(m)$.
• The corresponding bijection $\mathrm{U}(m)/\mathrm{O}(m) \to \Lambda(m)$ is continuous, hence, in view of the compactness of $\mathrm{U}(m)/\mathrm{O}(m)$, a homeomorphism. □

The given representation of $\Lambda(m)$ now permits us to define the Maslov index:

6.57 Lemma *The mapping* $MA_m : \Lambda(m) \to S^1$, $U\mathrm{O}(m) \mapsto \det\big(U\mathrm{O}(m)\big)^2$ *is well-defined.*

Proof: Regardless of the choice of $O \in \mathrm{O}(m)$, one has

$$\big[\det(UO)\big]^2 = \big[\det(U)\big]^2\big[\det(O)\big]^2 = \big[\det(U)\big]^2 .$$ □

The Maslov index of a loop in the Lagrange-Grassmann manifold $\Lambda(m)$ can also be calculated by assigning certain integers to its points of intersection with a certain subset $\overline{\Lambda_1}(m) \subset \Lambda(m)$, and adding these integers up.

6.58 Example (**Maslov Index for the 1D Harmonic Oscillator**) In Example 6.55.2, the Maslov index of the parametrization $\tilde{c} : S^1 \to H^{-1}(E)$ of the energy curve of a harmonic oscillator was calculated; more precisely, this was the Maslov index of the curve $c : S^1 \to \Lambda(1)$ associated with the curve $\tilde{c}$. Each point in $v \in \Lambda(1)$ gets passed over by c twice, namely in mathematically positive direction. So the Maslov index is

$$\deg(c) = \sum_{z \in c^{-1}(v)} \mathrm{sign}\big(c'(z)\big) = 2\,.$$ ◊

When generalizing this example to m degrees of freedom, it is often convenient to choose the 'vertical' Lagrangian subspace $v := \mathbb{R}^m_p \times \{0\} \subset \mathbb{R}^m_p \times \mathbb{R}^m_q$ as a reference point[17] $v \in \Lambda(m)$. Now for $m \geq 2$, typical smooth curves $c : S^1 \to \Lambda(m)$ will not hit the reference point because now $\dim\big(\Lambda(m)\big) = \frac{m(m+1)}{2} > 1$. However, if we consider the disjoint decomposition of the Lagrange-Grassmann manifold into

$$\Lambda(m) = \bigcup_{k=0}^{m} \Lambda_k(m) \quad \text{, with} \quad \Lambda_k(m) := \big\{u \in \Lambda(m) \mid \dim(u \cap v) = k\big\},$$

[17] Any other point of $\Lambda(m)$ could be chosen just as well.

then it will in general be unavoidable, as we'll see, for c to intersect

$$\overline{\Lambda}_1(m) := \Lambda(m) \backslash \Lambda_0(m)\,.$$

6.59 Exercises (**Maslov Index**)

1. Show that for all $k, m \in \mathbb{N}_0$, $k \le m$, the subset $\Lambda_k(m)$ is a fiber bundle (see Definition F.1) over the Grassmann manifold $\mathrm{Gr}(v, k)$ with the mapping

$$\pi_k : \Lambda_k(m) \to \mathrm{Gr}(v, k) \quad , \quad u \mapsto u \cap v$$

   in which the dimension of the fiber is

$$\dim\left(\pi_k^{-1}(w)\right) = \frac{(m-k+1)(m-k)}{2} \qquad \big(w \in \mathrm{Gr}(v, k)\big).$$

   Using Theorem 6.51, conclude that the manifold $\Lambda_k(m)$ has the dimension

$$\dim\big(\Lambda_k(m)\big) = \tfrac{1}{2}\big((m+1)m - (k+1)k\big) \qquad (k = 0, \ldots, m).$$

2. Show that $\overline{\Lambda}_1(m) = \overline{\Lambda_1(m)}$, hence for $k \ge 2$, the submanifolds $\Lambda_k(m)$ lie in the closure of $\Lambda_1(m)$, whereas $\Lambda_0(m)$ is open.

   Hint: Show first that $\pi_k^{-1}(w) \neq \emptyset$. Then consider, using Example 6.44.3, a neighborhood of $u \in \pi_k^{-1}(w)$. ♢

6.60 Example $(m = 1)$ $\Lambda_1(1) = \{v\}$, and $\Lambda_0(1) = \Lambda(1) \backslash \{v\}$ is diffeomorphic to $\mathbb{R}$. ♢

We already know that, generalizing this example, $\Lambda_0(m)$ is diffeomorphic to the vector space $\mathrm{Sym}(m, \mathbb{R})$, because precisely those Lagrangian subspaces that are transversal to the vertical subspace $V = \mathbb{R}_p^m \times \{0\}$ can be written as graphs $\{(Aq, q) \in \mathbb{R}_p^m \times \mathbb{R}_q^m\}$ of a symmetric $m \times m$ matrix A.

Therefore each loop $c : S^1 \to \Lambda_0(m)$, $c(z) = \mathrm{Graph}\big(A(z)\big)$ can be contracted to the constant loop (with value $\{0\} \times \mathbb{R}_q^m \in \Lambda_0(m)$) by means of the homotopy

$$H : S^1 \times [0, 1] \to \Lambda_0(m) \quad , \quad H(z, t) = \mathrm{Graph}\big(tA(z)\big)\,.$$

As the integer valued Maslov index cannot change under this continuous contraction, it is 0 for loops $c : S^1 \to \Lambda_0(m)$.

6.61 Exercise (**Range of the Maslov Index**)
For given $m \in \mathbb{N}$ and $I \in \mathbb{Z}$, there exists a curve $c : S^1 \to \Lambda(m)$ with

$$\deg\big(\mathrm{MA}_m \circ c\big) = I.$$

Hint: Using $\Lambda(1) \cong S^1$, first find such a curve for $m = 1$, then imbed it into $\Lambda(m)$ by viewing $\mathbb{R}_p \times \mathbb{R}_q$ as a symplectic subspace of $(\mathbb{R}^m_p \times \mathbb{R}^m_q, \omega_0)$. $\Diamond$

This exercise shows that intersections of $c(S^1)$ with $\overline{\Lambda}_1(m)$ are in general unavoidable. It is however possible to avoid, if need be by means of a homotopy, the set $\overline{\Lambda}_2(m) := \overline{\Lambda}_1(m) \backslash \Lambda_1(m)$, and to have, for $c(z) \in \Lambda_1(m)$, the direction vector $c'(z) \in T_{c(z)}\Lambda(m)$ transversal[18] to the tangent space $T_{c(z)}\Lambda_1(m)$; see the (schematic) Figure 6.5.1. This is due to the formula

$$\operatorname{codim}\big(\Lambda_k(m)\big) := \dim\big(\Lambda(m)\big) - \dim\big(\Lambda_k(m)\big) = \frac{(k+1)k}{2}$$

for the codimension of the submanifolds (see Exercise 6.59 and Figure 6.5.1). For $k \geq 2$, this codimension is at least 3, hence larger than $\dim(S^1) = 1$, and also larger than the dimension 2 of a loop moved by a homotopy $H : S^1 \times [0, 1] \to \Lambda(m)$ (which is why the following, second, definition of the Maslov index is independent of the choice of a homotopy).

If $c(z) \in \Lambda_1(m)$ and $c'(z)$ is transversal to $\Lambda_1(m)$, then exactly one eigenvalue $\lambda(z)$ of a unitary representation $U(z)$ of $c(z)$ is purely imaginary; and its derivative $\lambda'(z)$ has a well-defined orientation sign $\left(\frac{\lambda'(z)}{i\,\lambda(z)}\right)$. The sum of these signs equals the Maslov index of c.

6.62 Exercise (**Harmonic Oscillator**)
Calculate the Maslov index for the Lissajous figure from Figure 6.3.1 (right part), i.e., for a closed solution curve $\tilde{c} : S^1 \to F^{-1}(f) \subset \mathbb{R}^4$ with

$$F = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} \quad , \quad F_i(p, q) = \frac{\omega_i}{2}(p_i^2 + q_i^2) \quad \text{and} \quad \frac{\omega_2}{\omega_1} = \frac{5}{3}\,.$$

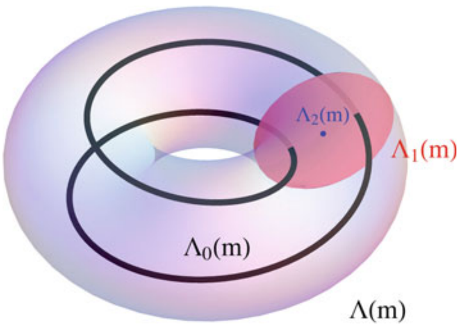


Figure 6.5.1 Schematic depiction of a closed curve $c : S^1 \to \Lambda(m)$ in the Lagrange-Grassmann manifold with Maslov index $\mathrm{MA}_m \circ c = 2$, and the submanifolds $\Lambda_k(m)$

[18]In HIRSCH [Hirs], one can find an extensive exposition on the subject of 'transversality'.

The mapping

$$c : S^1 \to \Lambda(2) \quad , \quad c(z) := T_{c(z)} F^{-1}(f)$$

is assigned to this curve. How can one determine, from the projection of $\tilde{c}$ onto the configuration space depicted in the figure, that $\deg(\mathrm{MA}_2 \circ c) = 16$? ◊

A Maslov index cannot only be assigned to loops of Lagrangian subspaces, but also to symplectic mappings. This is an immediate consequence of Theorem 6.49 in Chapter 6.4, because a loop $c : S^1 \to \mathrm{Sp}(2m, \mathbb{R})$ corresponds to a loop

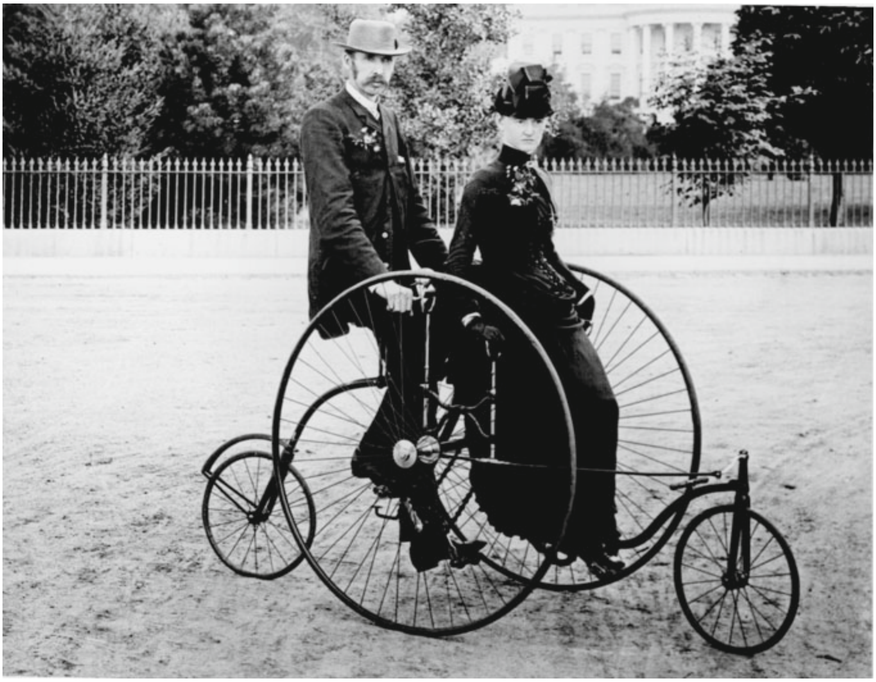
$$\tilde{c} : S^1 \to \Lambda(2m) \quad , \quad \tilde{c}(z) = \mathrm{graph}\big(c(z)\big) .$$

As a matter of fact, the *Maslov index of* c is defined as $\frac{1}{2}\deg(\mathrm{MA}_m \circ \tilde{c})$. The factor $\frac{1}{2}$ has the effect that every integer, rather than only even integers, can occur as a value of the Maslov index.

The Maslov index of $c : S^1 \to \mathrm{Sp}(2, \mathbb{R})$ can also be defined by means of the hypersurface of matrices with degenerate eigenvalues $+1$ (see Exercise 6.26).

6.63 Literature More advanced aspects are treated for example in the book [LiMa] by LIBERMANN and MARLE, and in the online manuscript 'Symplectic Geometry' by DUISTERMAAT [Dui] (which includes exercises). The monograph [Lon] by LONG has as its subject symplectic index theory. [SB] by SCHULZ- BALDES describes an alternative approach to calculating the Maslov index. ◊

Chapter 7
Stability Theory



1886 Coventry Rotary Quadracycle in Washington DC.[1]

The notions of stability introduced in Definition 2.21 have so far mainly been applied to linear systems. As we analyze nonlinear systems in the present chapter, asymptotic stability will turn out to be robust. An equilibrium is asymptotically

[1]Image: The U.S. National Archives and Records Administration.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_7

stable if the linearization of the vector field at this point is asymptotically stable (Theorem 7.6).

Unfortunately, asymptotic stability plays no role in Hamiltonian systems. Here instead, at least in the linear case, Lyapunov-stability transpires to be robust under small *Hamiltonian* perturbations. However, the proof of Lyapunov-stability in nonlinear Hamiltonian systems is often difficult; this subject will be resumed in Chapter 15.4 on KAM theory.

Bifurcations are topological changes of the phase space portrait of parameter dependent dynamical systems. In the simplest (*local*) case, they are a consequence of a change in stability properties of an equilibrium or a periodic orbit.

7.1 Stability of Linear Differential Equations

Before moving on from linear differential equations, let us study the qestion of their stability. Compared to the results from Chapter 5.2, the only new thing is the inclusion of non-hyperbolic matrices.

We again use the index of a square matrix, i.e., the sum of the algebraic multiplicities of those eigenvalues $\lambda \in \mathbb{C}$ that satisfy $\mathrm{Re}(\lambda) < 0$.

7.1 Theorem *The equilibrium* 0 *of the ODE* $\dot{x} = Ax$ *with* $A \in \mathrm{Mat}(n, \mathbb{R})$ *is*

1. **unstable** *if either* $\mathrm{Ind}(-A) > 0$*, or* $\mathrm{Ind}(-A) = 0$*, but there is an eigenvalue* $\lambda \in \mathrm{i}\,\mathbb{R}$ *whose algebraic multiplicity exceeds its geometric multiplicity;*
2. **Lyapunov-stable** *if neither of the conditions in part 1 are satisfied;*
3. **asymptotically stable** *if and only if* $\mathrm{Ind}(A) = n$*.*

Proof: As instability, Lyapunov-stability, and asymptotic stability of linear differential equations are not affected by similarity transformations, we may assume with no loss of generality that A is already in Jordan normal form, i.e., it consists of real Jordan blocks of the form

$$J_r^{\mathbb{R}}(\lambda) := J_r(\lambda) \text{ for } \lambda \in \mathbb{R} \quad \text{and} \quad J_r^{\mathbb{R}}(\lambda) := \begin{pmatrix} J_r(\mu) & -\varphi \mathbb{1}_r \\ \varphi \mathbb{1}_r & J_r(\mu) \end{pmatrix} \text{ for } \lambda \in \mathbb{C}\backslash\mathbb{R}\,,$$

with $\mu := \mathrm{Re}(\lambda)$, $\varphi := \Im(\lambda)$, and the Jordan blocks $J_r(\mu)$ from Definition 4.6.

1. By hypothesis, there exists a Jordan block for an eigenvalue λ, for which $\mathrm{Re}(\lambda) > 0$, or $\mathrm{Re}(\lambda) = 0$ but $r \geq 2$. We use as an initial value the rth canonical basis vector e_r in the flow-invariant subspace of this real Jordan block. Inspecting (4.1.6) or (4.1.7) respectively shows that

$$\lim_{t \to +\infty} \| \exp\bigl(J_r^{\mathbb{R}}(\lambda)\bigr) e_r \| = \infty\,.$$

2. Otherwise, there are no eigenvalues λ with real part $\mu > 0$, and those with real part $\mu = 0$ belong to Jordan blocks $J_r(\lambda)$ of size $r = 1$.

In the case $\mu < 0$, formula (4.1.6) for real λ and formula (4.1.7) for non-real λ show that

$$\lim_{t\to+\infty} \left\| \exp\left(J_r^{\mathbb{R}}(\lambda)t\right)\right\| = \lim_{t\to+\infty} e^{\mu t} \left\| \exp\left(J_r^{\mathbb{R}}(i\varphi)t\right)\right\| = 0\,.$$

For $\mathrm{Re}(\lambda) = 0$ and $r = 1$, apart from the case $\exp(J_r(0)t) = 1$, only the case of a real Jordan block

$$\left\|\exp\left(J_r^{\mathbb{R}}(\lambda)t\right)\right\| = \left\| \begin{pmatrix} \cos(\varphi t) & -\sin(\varphi t) \\ \sin(\varphi t) & \cos(\varphi t) \end{pmatrix} \right\| = 1$$

from (4.1.8) occurs; both cases lead to Lyapunov-stability.

3. If $\mathrm{Ind}(A) = n$, hence $\mathrm{Re}(\lambda) < 0$, then $\lim_{t\to\infty} \| \exp(J_r^{\mathbb{R}}(\lambda)t)\| = 0$.
   Otherwise there exists a λ with $\mathrm{Re}(\lambda) \geq 0$, hence

$$\lim_{t\to+\infty} \| \exp(J_r^{\mathbb{R}}(\lambda)t)\| \in \{1, \infty\}\,. \qquad \square$$

Hamiltonian flows do not have asymptotically stable fixed points because they leave the phase space volume invariant (this also applies to the nonlinear case, see Remark (10.14.2).

By Remark 6.25 and Theorem 7.1, the origin is a Lyapunov-stable fixed point of a linear Hamiltonian system if and only if all eigenvalues of the infinitesimally symplectic endomorphism generating the flow lie on the imaginary axis, and their algebraic and geometric multiplicities coincide.

So in a certain sense, Hamiltonian systems tend to have less stability than general dynamical systems. But there are other aspects as well.

7.2 Definition (Strong Stability for Linear Hamiltonian Systems)
An infinitesimally symplectic endomorphism $A \in \mathfrak{sp}(E, \omega)$ is called **strongly stable** *if there exists a neighborhood $U \subset \mathfrak{sp}(E, \omega)$ of A such that, for all $B \in U$, the origin is Lyapunov-stable with respect to the flow* $\exp(Bt)$.

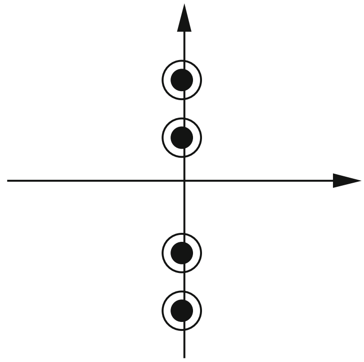
First notice in this definition that we are talking about stability not of a fixed point of a flow, but of an endomorphism generating a flow. This is of course owed to the fact that 0 is always a fixed point of a linear flow.

Moreover, this endomorphism is assumed to generate a Hamiltonian flow.

The underlying idea in the definition of strong stability is that in a particular mechanical system, we do not know the equation of motion with infinite precision, for example because we can only measure the masses of the interacting bodies with finite precision. We are therefore interested in the question if we can still decide whether the mechanical system is Lyapunov-stable.

7.3 Theorem (Strong Stability)
If all eigenvalues $\lambda \in \mathbb{C}$ of an infinitesimally symplectic endomorphism A are distinct and lie on the imaginary axis, then A is strongly stable.

Proof: Let the minimal distance of two zeros of the characteristic polynomial be 4ε, $\varepsilon > 0$. We consider for each zero an ε-neighborhood. These neighborhoods do not overlap, see the adjacent figure. The roots of the characteristic polynomial of a matrix depend continuously on the matrix entries. Thus an endomorphism B that is sufficiently close to A must have eigenvalues of which one each lies in one of the ε-neighborhoods.



Eigenvalues of a matrix $A \in \mathfrak{sp}(\mathbb{R}^4)$

If these eigenvalues λ did not at the same time lie on the imaginary axis, then $-\overline{\lambda} \neq \lambda$ would lie in the same ε-neighborhood; contradiction! □

7.4 Exercise (Strong Stability)
We consider the single linear, time dependent differential equation

$$\ddot{x}(t) = -f(t)\,x(t) \qquad (t \in \mathbb{R})$$

with continuous, T-periodic $f : \mathbb{R} \to [0, \infty)$, $T > 0$, hence $f(t + T) = f(t)$. This differential equation is equivalent to a Hamiltonian system with time dependent Hamiltonian

$$H : \mathbb{R}_t \times \mathbb{R}^2_x \to \mathbb{R} \quad , \quad H(t, x_1, x_2) := \tfrac{1}{2}x_1^2 + \tfrac{1}{2} f(t)\, x_2^2 \,.$$

(a) Show that the linear mappings $\Phi_t : \mathbb{R}^2 \to \mathbb{R}^2$, $\Phi_t(x(0)) := x(t)$, $t \in \mathbb{R}$, defined by the Hamiltonian flow do not in general form a group, due to the time dependence of f.
(b) Show that $\Phi_{T+s} = \Phi_s \circ \Phi_T$ for all $s \in \mathbb{R}$. It then follows that the mappings $\Phi_{nT} = (\Phi_T)^n$, $n \in \mathbb{Z}$, form a group.
(c) Then $A := \Phi_T : \mathbb{R}^2 \to \mathbb{R}^2$ is called the *period map* of the system. Explain why this map is linear and has determinant 1.
(d) Show that the zero solution $\Phi_t(0) = 0$ is Lyapunov-stable if and only if 0 is a Lyapunov-stable fixed point of A, i.e., of the dynamical system

$$\Psi : \mathbb{Z} \times \mathbb{R}^2 \to \mathbb{R}^2 \quad , \quad \Psi(n, x) := A^n(x)\,.$$

(e) Generalizing Definition 7.2, we call the zero solution *strongly stable*, if it is Lyapunov-stable for all Hamiltonians in a neighborhood of H. Here we limit ourselves to Hamiltonians of the form $\tilde{H} : \mathbb{R}_t \times \mathbb{R}^2 \to \mathbb{R}$, $(t, x) \mapsto \langle x, \tilde{B}(t)x \rangle$ with continuous $\tilde{B} : \mathbb{R} \to \mathrm{Mat}(2, \mathbb{R})$, $\tilde{B}(t) = \tilde{B}(t)^\top = \tilde{B}(t + T)$ and use the norm

$$\|\tilde{H}\| := \sup\left\{\|B(t)v\|_{\mathbb{R}^2} \mid t \in [0,T],\ v \in \mathbb{R}^2,\ \|v\| \le 1\right\}$$

to define a neighborhood. Show that the zero solution is strongly stable if A (from (c)) satisfies the condition $|\mathrm{tr}\,(A)| < 2$.

(f) Now consider in particular the function $f(t)^2 := \omega^2(1+\varepsilon\cos(t))$ with $\omega \in \mathbb{R}$ and $|\varepsilon| < 1$. Show by means of (e) that the zero solution for the parameters $\varepsilon = 0$ and $\omega \in \mathbb{R} \setminus \frac{1}{2}\mathbb{Z}$ is strongly stable. ◊

7.2 Lyapunov Functions

We now leave the case of linear systems of differential equations and turn to the nonlinear case.

A natural question is whether a fixed point is asymptotically stable, provided the linearization of the ODE about this fixed point leads to an asymptotically stable linear flow. Lyapunov functions are useful to answer this question and similar ones (like the question addressed in Remark 7.9).

7.5 Definition *Let* $\Phi : G \times M \to M$ *be a continuous dynamical system.*[2] $L \in C^0(M,\mathbb{R})$ *is called a* **Lyapunov function** *for* Φ*, if* L *(oder* $-L$*) decreases along orbit curves.*

Similar to the case discussed in Section 6.1.1 on gradient systems, an orbit curve that, at a given time, meets the sublevel set $M_\ell := \{m \in M \mid L(m) \le \ell\}$, cannot leave this set in the future. On the other hand, one would like for the function to decrease *strictly* in time (other than at fixed points), so that the trajectories are trapped in smaller and smaller sets $M_{\ell'} \subset M_\ell$ where $\ell' < \ell$.

Of course, a corresponding argument can also be used for functions that are increasing along orbit curves.

We now want to show that the solutions of the differential equation are controlled, in leading order, by the linearization of f, as long as we are close to an equilibrium. We can use Gronwall's inequality (Theorem 3.42).

7.6 Theorem (Lyapunov)
An equilibrium $x_s \in U \subseteq \mathbb{R}^n$ *of the differential equation*

$$\dot{x} = f(x) \quad , \quad f \in C^1(U,\mathbb{R}^n)$$

is asymptotically stable if $\mathrm{Ind}\big(\mathrm{D}f(x_s)\big) = n$.

---

[2]When using Lyapunov functions, it often suffices to show that the dynamics $\Phi_t : M \to M$ exists for all $t \ge 0$, as is the case for example in the proof of the following Theorem 7.6.

Proof:

- By shifting coordinates, we may again assume that $x_s = 0$. Since $U \subseteq \mathbb{R}^n$ is open, a ball of some positive radius $\tilde{r}$ belongs to U.

- For $A := \mathrm{D}f(x_s)$, there exists $\Lambda < 0$ with $\mathrm{Re}(\lambda_i) < \Lambda$ for all eigenvalues $\lambda_i \in \mathbb{C}$ of A. Moreover there exist some $C \geq 1$ (as can be read off the Jordan normal form of $\exp(At)$) with
$$\|\exp(At)\| \leq C\exp(\Lambda t) \qquad (t \geq 0). \tag{7.2.1}$$
The Duhamel equation (4.2.10) permits to write the maximal solution to the initial value problem in the form
$$x(t) = \exp(At)x(0) + \int_0^t \exp\big(A(t-s)\big)R\big(x(s)\big)\,\mathrm{d}s \qquad (t \in I)$$
with $I \subseteq \mathbb{R}$, with the source term $R(x) := f(x) - Ax$, evaluated along $s \mapsto x(s)$.

- Now there exists some radius $r \in (0, \tilde{r})$ such that
$$\|R(x)\| \leq \frac{|\Lambda|}{2C}\|x\| \qquad \text{if } \|x\| \leq r\,, \tag{7.2.2}$$
because by Taylor, $\lim_{x\to 0}\frac{\|R(x)\|}{\|x\|} = \lim_{x\to 0}\|f(x) - \mathrm{D}f(0)x - f(0)\|/\|x\| = 0$.

- Using (7.2.1) and (7.2.2), it follows for the function
$$F : I \to [0,\infty) \quad , \quad F(t) := \|x(t)\|\exp(|\Lambda|t)\,,$$
for all solutions with initial values $x(0) \in U_r(0)$, and for the maximal time interval $[0, T) \subset I$ with $x\big([0,T)\big) \subset U_r(0)$ that
$$F(t) \leq CF(0) + \int_0^t C\frac{|\Lambda|}{2C}\|x(s)\|\,\mathrm{d}s \leq CF(0) + \int_0^t \frac{|\Lambda|}{2}F(s)\,\mathrm{d}s \quad \big(t \in [0,T)\big).$$
By Gronwall's lemma 3.42, one concludes $F(t) \leq CF(0)\exp(\frac{1}{2}|\Lambda|t)$, or
$$\|x(t)\| \leq C\|x(0)\|\exp\left(\tfrac{1}{2}\Lambda t\right).$$
Thus the solution curves with initial condition $x(0) \in U_{r/C}(0)$ stay in the ball $U_r(0)$ for all positive times and converge to the equilibrium at 0. In other words, x_s is asymptotically stable. □

7.7 Remarks

1. The proof also provided the statement that all $x \in U_{r/C}(0)$ belong to orbits that converge to the equilibrium, i.e., are in its *basin*, (see Def. on page 20).
2. Although the function $x \mapsto \|x\|$ used in the proof of the theorem is in general not a Lyapunov function, it served the same purpose, as $\lim_{t\to+\infty} F(t) = 0$.
3. As a corollary, we obtain that an equilibrium x_s is unstable if $\mathrm{Ind}\big(-\mathrm{D}f(x_s)\big) = n$. This follows from Theorem 7.6 by time reversal, i.e., by transition to $-f$. However, for the instability of x_s it already suffices that $\mathrm{Ind}\big(-\mathrm{D}f(x_s)\big) \geq 1$, that is, there is one eigenvalue λ with $\mathrm{Re}(\lambda) > 0$ (see e.g. [Wa1], §29). $\Diamond$

7.8 Exercise (Lyapunov Function)
The vector fields in the following exercise are not complete. So we extend definition 7.5 to this situation: We assume that for the differential equation $\dot{x} = g(x)$ on $M \subseteq \mathbb{R}^n$ with a locally Lipschitz continuous vector field $g : M \to \mathbb{R}^n$, there exists a continuously differentiable function $V : M \to \mathbb{R}$, such that all solutions $I \ni t \mapsto x(t)$ to the differential equation satisfy

$$\tfrac{\mathrm{d}}{\mathrm{d}t} V\big(x(t)\big) \leq 0 \qquad (t \in I).$$

As we know, the origin is a Lyapunov-stable, but not asymptotically stable, fixed point of the linear system $\dot{x} = Ax$ with $x \in \mathbb{R}^2$, $A = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

Use its Lyapunov function $V(x) := \frac{1}{2}\|x\|^2$ to investigate the stability of the perturbed systems $\dot{x} = Ax + f_j(x)$ with $f_j : \mathbb{R}^2 \to \mathbb{R}^2$ ($j \in \{1, 2, 3\}$):

(a) $f_1(x_1, x_2) := (-x_1^3 - x_1 x_2^2,\ -x_2^3 - x_1^2 x_2)^\top$
(b) $f_2(x_1, x_2) := (x_1^3 + x_1 x_2^2,\ x_2^3 + x_1^2 x_2)^\top$
(c) $f_3(x_1, x_2) := (-x_1 x_2,\ x_1^2)^\top$.

Which shape do the orbits have in case (c)?

Now let $\dot{x} = f_4(x) := \begin{pmatrix} -x_2 - x_1 x_2^2 + x_3^2 - x_1^3 \\ x_1 + x_3^3 - x_2^3 \\ -x_1 x_3 - x_1^2 x_3 - x_2 x_3^2 - x_3^5 \end{pmatrix}$. Show:

(d) $0 \in \mathbb{R}^3$ is asymptotically stable.
(e) The trajectories of the linearized system $\dot{x} = \mathrm{D}f_4(0)x$ lie on circles parallel to the x_1, x_2 plane, so for the linearized system, the origin is Lyapunov-stable, but not asymptotically stable. $\Diamond$

7.9 Remark (Lyapunov Function and Hamiltonian Dynamics)
It may appear that the notion of a Lyapunov function does not apply in the Hamiltonian case. Indeed, the phase space volume is conserved in this case, so asymptotic stability cannot be expected. There are however other applications:

1. Nondegenerate extrema x_s of a Hamilton function H are Lyapunov-stable fixed points, because on one hand, $X_H(x_s) = 0$, on the other hand, H itself is a Lyapunov function since the value of H does not change along orbits. Thirdly,

for a minimum x_s, every neighborhood U of x has a neighborhood $V \subset U$ of x_s in the form $V = H^{-1}\big([h, h+\varepsilon)\big)$, with $h := H(x_s)$.

2. As a different application, consider the free motion $\dot{q} = p$, $\dot{p} = 0$ with the flow $\Phi\big(t, (p_0, q_0)\big) = (p_0, q_0 + p_0 t)$, given by the Hamiltonian

$$H : \mathbb{R}^d_p \times \mathbb{R}^d_q \to \mathbb{R} \quad , \quad H(p, q) = \tfrac{1}{2}\|p\|^2 .$$

Then $L : \mathbb{R}^d_p \times \mathbb{R}^d_q \to \mathbb{R},\ (p, q) \mapsto \langle p, q\rangle$ is a Lyapunov function because

$$\frac{\mathrm{d}}{\mathrm{d}t} L\big(\Phi_t(p, q)\big) = \langle \dot{p}, q\rangle + \langle p, \dot{q}\rangle = \langle p, \dot{q}\rangle = \|p\|^2 \geq 0 .$$

Since $L(p, q) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|q\|^2$, the function $t \mapsto \|q\|^2(t)$ is convex. In this example, we could have calculated this directly, because $\|q\|^2(t) = \|q_0 + p_0 t\|^2$.
But we can argue similarly if the motion of the particle is perturbed weakly by a force field. Under appropriate hypotheses, we can then conclude that the trajectory still goes to spatial infinity as time $t \to \pm\infty$. This is why this argument is popular in scattering theory (see for instance the proof of Theorem 12.5). In that context, L is called *escape function*. ◊

7.3 Bifurcations

A dynamical system will frequently depend on parameters, like the example of a spring with friction discussed in Chapter 5.4. As the parameter is changed, in general, the phase portrait (i.e., the decomposition of phase space into orbits) will change.

It is possible that the dynamical systems for any two values of the parameters are conjugate in the sense of Definition 2.28, so that we can find a homeomorphism of the phase space that maps the phase portraits into each other, i.e., that maps oriented orbits onto oriented orbits. We constructed such homeomorphisms for the linear hyperbolic flows with the same index in Theorem 5.9.

But it can also occur that for a particular choice of the parameter, the phase portrait changes qualitatively, in other words that it is not merely a deformation of the phase portrait for other parameters, and thus such a homeomorphism does not exist. Such a phenomenon is called *bifurcation*. The notion goes back to Poincaré, as do many things in the theory of dynamical systems.

Qualitative changes of the phase portrait by bifurcation of a fixed point are called *local* bifurcations, all others *global* bifurcations.

A special situation arises if the phase space itself changes in dependence on a parameter. For instance, in Hamiltonian systems, the total energy is a constant of motion, so one can use its level surfaces as reduced phase spaces. As the energy changes, these phase spaces can either remain homeomorphic to each other, or they can change in shape.

7.3.1 Bifurcations from Equilibria

The simplest case is the theory of local bifurcations of a parameter dependent differentiable dynamical system. Its phase space M and the parameter space P are assumed to be differentiable manifolds.
So we study, for a level of differentiability $n \in \mathbb{N}$ (or $n = \infty$):

- in case of *discrete time* $\mathbb{Z}$, a parameter dependent diffeomorphism
$$F^{(p)} \equiv F(\cdot, p) : M \to M \quad (p \in P)$$
given by some $F \in C^n\big(M \times P, M\big)$, and a fixed point $m_0 \in M$ of $F^{(p_0)}$;
- in case of *continuous time* $\mathbb{R}$, the parameter dependent differential equation
$$\dot{m} = f^{(p)}(m) \quad \text{for} \quad f^{(p)}(m) := f(m, p) \text{ and } f \in C^n\big(M \times P, TM\big), \tag{7.3.1}$$
with the tangent bundle $\pi_M : TM \to M$ of the manifold M (see Appendix A.3) and vector fields $f^{(p)}$ on M (that is, $\pi_M \circ f^{(p)} = \mathrm{Id}_M$). For the parameter $p_0 \in P$, let the point $m_0 \in M$ in phase space be an equilibrium ($f^{(p_0)}(m_0) = 0$).

For local considerations in a neighborhood of $(m_0, p_0) \in M \times P$, we assume with no loss of generality that the phase space M is an open subset of $\mathbb{R}^d$, rather than an arbitrary manifold, and that the parameter space P is open in $\mathbb{R}^k$. Then we can consider the vector field as a map $f : M \times P \to \mathbb{R}^d$.

7.10 Definition *$p_0 \in P$ is called* **bifurcation point for the equilibrium** *m_0 if*

- $\mathrm{D}_1 F(m_0, p_0) \in \mathrm{Mat}(d, \mathbb{R})$ *has a complex eigenvalue with modulus 1;*
- $\mathrm{D}_1 f(m_0, p_0) \in \mathrm{Mat}(d, \mathbb{R})$ *has an imaginary eigenvalue.*

7.11 Remark
On one hand, this definition has the advantage that the condition can be checked easily. On the other hand, we know as a consequence of Theorem 5.9 that for *linear* differential equations $\dot{x} = Ax$, these are exactly the non-hyperbolic system matrices $A \in \mathrm{Mat}(d, \mathbb{R})$ for which bifurcations do occur. And it is exactly in this case that the solution operators $\exp(At)$ for times $t \neq 0$ have a complex eigenvalue of modulus 1; this motivates the first part of the definition. ♢

At these points $(m_0, p_0) \in M \times P$, the parameter dependent phase portrait *can* change qualitatively, but it doesn't have to. Conversely, however, one has

7.12 Theorem (Parametrized fixed points) *Let $m_0 \in M$ be a fixed point for the parameter p_0. Under the following hypotheses, there exist open neighborhoods $\tilde{P} \subseteq P$ of p_0 and $\tilde{M} \subseteq M$ of m_0 and a mapping $\widehat{m} \in C^n(\tilde{P}, \tilde{M})$ with $\widehat{m}(p_0) = m_0$ that parametrizes the equilibria locally:*

- *If 1 is not an eigenvalue of the matrix $\mathrm{D}_1 F(m_0, p_0)$. Then for $p \in \tilde{P}$,*

$$F\big(\widehat{m}(p), p\big) = \widehat{m}(p) \quad , \textit{and} \quad \mathrm{graph}(\widehat{m}) = \{(p, m) \mid F(m, p) = m\} \cap \tilde{M} \times \tilde{P}\,.$$

- *If 0 is not an eigenvalue of the matrix $\mathrm{D}_1 f(x_0, p_0)$. Then*

$$f\big(\widehat{m}(p), p\big) = 0 \quad , \textit{and} \quad \mathrm{graph}(\widehat{m}) = \{(p, m) \mid f(m, p) = 0\} \cap \tilde{M} \times \tilde{P}\,.$$

Proof: This is simply the implicit function theorem, applied in a chart of $P \times M$ to $F - \mathrm{pr}_M$ with $\mathrm{pr}_M : M \times P \to M$, $(m, p) \mapsto m$, or to f, respectively. □

In each case, the hypothesis is satisfied if m_0 is not a bifurcation point.

7.13 Remark (Asymptotic Stability)
The theorem says that under small changes of the parameters, the fixed point cannot disappear, nor can a second fixed point appear in its proximity. Instead, its location changes according to $\widehat{m}$, namely with as much differentiability as the parameter dependent diffeomorphism F or vector field f itself has.

If one even requires that p_0 is not a bifurcation point for the equilibrium m_0, then in the case of continuous time, the index $p \mapsto \mathrm{Ind}\big(\mathrm{D}_1 f(\widehat{m}(p), p)\big)$ of the system matrix (Definition 5.2) is constant in a neighborhood of p_0.
A similar statement applies to the case of discrete time, for the number of complex eigenvalues with modulus strictly less than 1. ◊

The situation is different in the following examples. The fact that they don't generate dynamical systems in the narrow sense since the solutions do not exist for all times is of no relevance to the issue.

7.14 Example (Saddle-Node Bifurcation)
We consider the family of differential equations $\dot{x} = f^{(p)}(x) := p + \frac{1}{2}x^2$, parametrized by $p \in P := \mathbb{R}$, on the phase space $M := \mathbb{R}$.

- For $p > 0$, one has $f^{(p)}(x) \geq p > 0$, so there exists no fixed point.
- For $p = 0$, there exists only the fixed point $x_0 = 0$. It is unstable because the general solution $x(t) = \frac{x_0}{1 - t\, x_0/2}$ even diverges for the initial values $x_0 > 0$, namely at time $t = 2/x_0$.
- For $p < 0$, there exist two fixed points $x_\pm(p) := \pm\sqrt{-2p}$. $x_-(p)$ is an asymptotically stable, and $x_+(p)$ an unstable equilibrium, as $Df^{(p)}(x_\pm(p)) = x_\pm(p)$.

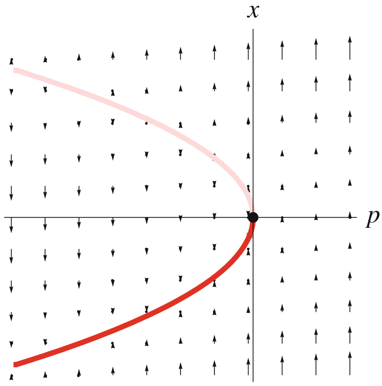
The linearization of the differential equation at the fixed points is of the form

$$\dot{y} = x_{\pm}(p) \cdot y \quad \text{for} \quad p < 0$$

and

$$\dot{y} = 0 \quad \text{for} \quad p = 0\,.$$

So it is exactly at $p = 0$ that the linearization doesn't have the maximal rank 1.



Saddle-node bifurcation $\dot{x} = p + \frac{1}{2}x^2$.
unstable fixed point: light red

This example shows how a pair consisting of a stable and an unstable fixed point can disappear when changing a parameter p. ♢

7.15 Example (Hopf Bifurcation) On the phase space $M := \mathbb{R}^2$, the normal form of the Hopf DE is

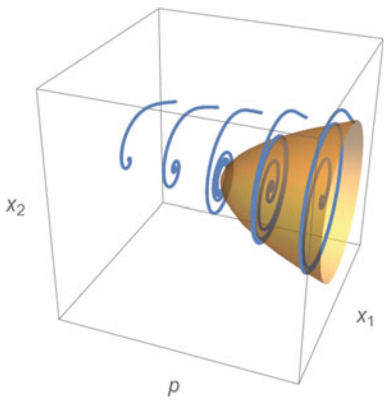
$$\dot{x} = \begin{pmatrix} p & -1 \\ 1 & p \end{pmatrix} x - \|x\|^2 x\,, \tag{7.3.2}$$

with parameter $p \in \mathbb{R}$, see the adjacent figure. Obviously, $0 \in M$ is an equilibrium for all values of the parameter. It is also the only equilibrium since the norm squared of the vector field on the right side of (7.3.2) is, for $x \neq 0$,

$$\|\dot{x}\|^2 = \|x\|^2 \left(1 + (p - \|x\|^2)^2\right) > 0.$$

At the only bifurcation point $p_0 = 0$ for this equilibrium, one has $Df^{(0)}(0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$,

so the eigenvalues are not real. So the equilibrium is preserved, but its stability changes: On $\mathbb{R}^2 \setminus \{0\}$, the ODE in polar coordinates, $x_1 = r\cos\varphi$, $x_2 = r\sin\varphi$, $(r, \varphi) \in (0, \infty) \times \mathbb{R}$, is of the form



The Hopf bifurcation (7.3.2)

$$\dot{r} = r\left(p - r^2\right) \quad , \quad \dot{\varphi} = 1\,,$$

because $\dot{r} = \frac{\frac{d}{dt} r^2}{2r} = \frac{x_1\dot{x}_1 + x_2\dot{x}_2}{r}$ and $\dot{\varphi} = \frac{\dot{x}_2 x_1 - \dot{x}_1 x_2}{x_1^2 + x_2^2} = 1$.

So the time derivative $\dot{r}$ will be negative when $r^2 > \max(0, p)$, and positive when $0 < r^2 < p$. For $r^2 = p > 0$, one has $\dot{r} = 0$, and we have found a periodic orbit of period 2π (see also Example 3.35.2). $\diamondsuit$

In a sense, these two examples are typical for one-parameter bifurcations from fixed points. Let us consider a continuous family

$$A : I \to \mathrm{Mat}(d, \mathbb{R}) \quad , \quad t \mapsto A_t$$

of matrices parametrized over an interval I. We assume that for parameter $t = 0 \in I$, the matrix A_0 has at least one eigenvalue λ_0 with real part 0. Then there are two cases: $\lambda_0 = 0$ or $\lambda_0 \neq 0$. In the second case, $\lambda_0, \overline{\lambda_0} \in \mathrm{i}\,\mathbb{R} \setminus \{0\}$ are two distinct imaginary eigenvalues. If A_0 has maximal rank subject to this stipulation, then a real Jordan normal form of A_0 is

– for $\lambda_0 = 0$, equal to $0 \oplus J \in \mathbb{R} \oplus \mathrm{Mat}(d-1, \mathbb{R})$
– for $\lambda_0, \overline{\lambda_0} \in \mathrm{i}\mathbb{R} \setminus \{0\}$, equal to $\begin{pmatrix} 0 & -\Im(\lambda_0) \\ \Im(\lambda_0) & 0 \end{pmatrix} \oplus J \in \mathrm{Mat}(2, \mathbb{R}) \oplus \mathrm{Mat}(d-2, \mathbb{R})$,

where in both cases the matrix J is regular. The first case occurs in a saddle-node bifurcation, the second in a Hopf bifurcation.

7.3.2 Bifurcations from Periodic Orbits

As for bifurcation from a t_0-periodic orbit $\mathcal{O}(m_0)$ in a parameter dependent dynamical system $\Phi^{(p)} : G \times M \to M$:

- In the case of *discrete time* $G = \mathbb{Z}$ and $\Phi_1^{(p)} = F^{(p)} : M \to M$, it can be reduced to bifurcation from the fixed point m_0 of the iterate $\Phi_{t_0}^{(p)} : M \to M$.
- In the case of *continuous time* $G = \mathbb{R}$ and $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t^{(p)}\big|_{t=0} = f^{(p)}$, the situation can *also* be reduced to bifurcation of the fixed point m_0 of a certain mapping, which is called the Poincaré map.

7.16 Definition

- *For a vector field $X \in C^n(M, TM)$ on the manifold M, a submanifold S of* $\dim(S) = \dim(M) - 1$ *will be called a* **(local) section transversal** *to X, if*

$$T_m M = T_m S \oplus \mathrm{span}(X(m)) \qquad (m \in S).$$

- *For the differentiable dynamical system $\Phi \in C^n(\mathbb{R} \times M, M)$ of the vector field $X = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_t|_{t=0}$ and a point $m \in M$ on a periodic orbit, let $S \subset M$ be a section transversal to X, with $m \in S$. For open neighborhoods $U, V \subseteq S$ of m, a diffeomorphism $F : U \to V$ is called* **Poincaré map** *of the orbit $\mathcal{O}(m)$, if F is of the form*

$$F(x) = \Phi\big(T(x), x\big) ,$$

with the **Poincaré time**

$$T : U \to (0,\infty) \quad , \quad T(x) := \inf\{t > 0 \mid \Phi_t(x) \in V\} .$$

7.17 Theorem
For every periodic orbit $\mathcal{O}(m)$ of $\Phi \in C^n(\mathbb{R}\times M, M)$ with minimal period $t_0 > 0$, there exist

- *a section $S \subset M$ with $m \in S$, transversal to the vector field X, and*
- *a Poincaré map $F \in C^n(U, V)$ with open neighborhoods $U, V \subseteq S$ of m, whose Poincaré time $T \in C^n(U, \mathbb{R}^+)$ satisfies $T(m) = t_0$.*

Proof:

- Since m is not an equilibrium, hence the vector field $X = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_t|_{t=0}$ is nonzero at m, there exists a transversal section at m: Namely, by the straightening theorem (Theorem 3.46), there exists a C^n coordinate chart (W, φ) of M with $m \in W$ that maps $X\restriction_W$ onto the constant vector field $e_1 = (1, 0, \ldots, 0)^\top$ on $\varphi(W) \subseteq \mathbb{R}^n$. Here $d := \dim(M)$.
  With no loss of generality, we simply set $\varphi(m) := 0 \in \mathbb{R}^d$ and calculate in the local coordinates of the chart. With the projections

$$\Pi : \mathbb{R}^d \to \mathbb{R}^d,\ \Pi(x) := \langle x, e_1\rangle\, e_1 = \varphi_1(x)e_1 \quad \text{and} \quad \Pi^\perp := 1\!\!1_d - \Pi , \tag{7.3.3}$$

  the set

$$W_\varepsilon := \{x \in W \mid \|\Pi(x)\| < \varepsilon,\ \|\Pi^\perp(x)\| < \varepsilon\}$$

  is a cylinder for small $\varepsilon > 0$, with $W_\varepsilon = (-\varepsilon, \varepsilon) \times S_\varepsilon$ for

$$S_\varepsilon := \{x \in W_\varepsilon \mid \Pi(x) = 0\} .$$

  On W_ε, the flow is of the form

$$\Phi_t(x) = x + t\, e_1 \qquad (x \in S_\varepsilon,\ |t| < \varepsilon), \tag{7.3.4}$$

  because $X = e_1$ on W.

- We let $U := \{x \in S_\varepsilon \mid \Phi_{t_0}(x) \in W_\varepsilon\}$. Then with

$$T : U \to (0,\infty) \quad , \quad T(x) := t_0 - \varphi_1\big(\Phi_{t_0}(x)\big)$$

  and $F : U \to V := F(U),\ F(x) := \Phi\big(T(x), x\big)$, we have

$$F(x) = \Phi\big(-\varphi_1(\Phi_{t_0}(x)),\ \Phi_{t_0}(x)\big) \in S_\varepsilon ,$$

because (7.3.4) implies $\Phi\big(-\varphi_1(y), y\big) \in S_\varepsilon$ for all $y \in W_\varepsilon$. Since the flow Φ is in $C^n(\mathbb{R} \times M, M)$ and the coordinate φ_1 is in $C^n(W, \mathbb{R})$, it follows that $T \in C^n(U, \mathbb{R}^+)$ and $F \in C^n(U, V)$.

- For $\varepsilon > 0$, sufficiently small, T is a Poincaré time, i.e., $T = \tilde{T}$ with

$$\tilde{T} : U \to \mathbb{R}^+ \quad , \quad \tilde{T}(x) := \inf\{t > 0 \mid \Phi_t(x) \in S_\varepsilon\} .$$

because for $t \in (\varepsilon, t_0 - \varepsilon)$, it follows that $\Phi_t(m) \notin W_\varepsilon$, else t_0 would not be the minimal period. On the other hand, $|T(x) - t_0| < \varepsilon$ for all $x \in U$, so we obtain a contradiction to the existence of a sequence of $x_n \in U$ with $\lim_{n\to\infty} x_n = m$ and $\lim_{n\to\infty} T(x_n) \in (0, t_0)$. □

7.18 Remark (Benefit of the Poincaré Map)
Whereas in general, the Poincaré map $F : U \to V$ from Theorem 7.17 does not define a discrete dynamical system, since domain and range do not coincide, the information contained in F is nevertheless sufficient to analyze bifurcations from the periodic orbit. It is actually more appropriate for the purpose than the mapping $\Phi_{t_0} : M \to M$, which has m as a fixed point as well.
Here is why: The linear mapping $T_m\Phi_{t_0}$ on the tangent space T_mM has $X(m)$ as an eigenvector for eigenvalue 1:

$$T_m\Phi_{t_0}\big(X(m)\big) = T_m\Phi_{t_0}\left(\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t(m)\Big|_{t=0}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_{t_0+t}(m)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_t(m)\Big|_{t=0},$$

which equals $X(m)$. In other words, the linear mapping $T_m\Phi_{t_0} - \mathrm{Id}_{T_mM}$ has on T_mM a nontrivial kernel. So in the case of a parameter dependent dynamical system

$$\Phi^{(p)} : \mathbb{R} \times M \to M \qquad (p \in P),$$

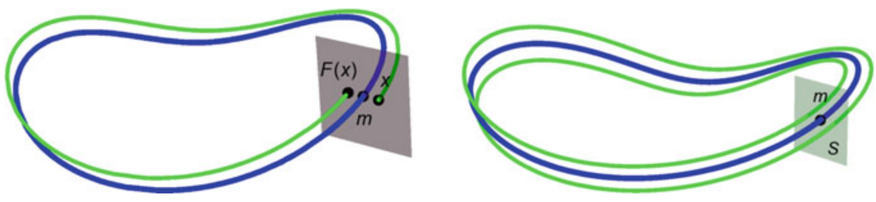


Figure 7.3.1 Left: Poincaré section; right: Bifurcation of a periodic orbit

if $\Phi^{(p_0)}$ has a periodic point m_0 with period t_0, one cannot use the implicit function theorem to conclude that $\Phi^{(p)}$ has a periodic point $\widehat{m}(p)$ of the same period t_0 for p in a neighborhood of p_0. And usually this is not the case.
However, typically there *do exist* periodic points $\widehat{m}(p)$ of $\Phi^{(p)}$ with minimal period $t(p)$, $m(p_0) = m_0$, $t(p_0) = t_0$, and continuous maps $p \mapsto \widehat{m}(p)$, $p \mapsto t(p)$. ◇

7.19 Lemma (Eigenvalues of the Linearized Poincaré Map) *If $\mathcal{O}(m) \subseteq M$ is a t_0-periodic orbit of the flow $\Phi : \mathbb{R} \times M \to M$, and if $F : U \to V$ is a Poincaré map of this orbit with $F(m) = m$, then the complex eigenvalues satisfy*[3]

$$\operatorname{spec}\big(D\Phi_{t_0}(m)\big) = \operatorname{spec}\big(DF(m)\big) \cup \{1\}\,.$$

Proof: As the transversal section $S_\varepsilon \subset M$ containing U and V has the property $T_m M = T_m S_\varepsilon \oplus \operatorname{span}\big(X(m)\big)$ according to Definition 7.16, we can choose adjusted coordinates just as in the proof of Theorem 7.17.
In these, $x \in W_\varepsilon = (-\varepsilon, \varepsilon) \times S_\varepsilon$ is of the form

$$x = y + s\,e_1 \quad \text{with} \quad (s\,e_1, y) := \big(\Pi(x), \Pi^\perp(x)\big) \quad \text{, hence} \quad s \in (-\varepsilon, \varepsilon),\; y \in S_\varepsilon$$

with the projection Π from (7.3.3). With representation (7.3.4) of the flow in the chart, $\Phi_{t_0}(y + s\,e_1) = \Phi_{t_0+s}(y)$ equals

$$\Phi_{s+t_0-T(y)} \circ \Phi_{T(y)}(y) = \Phi_{s+t_0-T(y)} \circ F(y) = F(y) + \big(s + t_0 - T(y)\big)e_1\,,$$

thus

$$D\Phi_{t_0}(m)\begin{pmatrix}\delta s\\ \delta y\end{pmatrix} = \begin{pmatrix}1 & -DT(m)\\ 0 & DF(m)\end{pmatrix}\begin{pmatrix}\delta s\\ \delta y\end{pmatrix}.$$

This block structure implies the claim about the eigenvalues. □
We have even proved that the algebraic multiplicity of the eigenvalue 1 of $D\Phi_{t_0}(m)$ is by one larger than that of $DF(m)$. From this, we obtain

7.20 Corollary (Parametrized Periodic Orbits)
For the parameter dependent flow $\Phi^{(p)} : \mathbb{R} \times M \to M$ $(p \in P)$ of (7.3.1), let $\mathcal{O}(m_0) \subseteq M$ be a periodic orbit of $\Phi^{(p_0)}$ with minimal period $t_0 > 0$.
If the eigenvalue 1 of $T_{m_0}\Phi^{(p_0)}_{t_0}$ has only algebraic multiplicity one, then this periodic orbit may be continued in the following sense:
There exist a neighborhood $\tilde{P} \subseteq P$ of p_0 in the parameter space and mappings $\widehat{m} \in C^n(\tilde{P}, S_\varepsilon)$ and $t \in C^n\big(\tilde{P}, (0, \infty)\big)$, with $\widehat{m}(p_0) = m_0$ and $t(p_0) = t_0$, such that the $\Phi^{(p)}$-orbit through $\widehat{m}(p)$ is periodic with minimal period $t(p)$.

7.21 Exercise (Parametrized Periodic Orbits) Prove Corollary 7.20. Make precise in which sense these orbits are essentially unique. ◊

7.22 Remark (Bifurcations from Periodic Orbits)
If the eigenvalue 1 has a multiplicity larger than one, further periodic orbits with similar period could bifurcate from the given periodic orbit.

The latter doesn't happen if the hypotheses of Corollary 7.20 are satisfied, but other bifurcation phenomena could still occur. For instance, as indicated in the right

---

[3]The **spectrum** $\operatorname{spec}(A)$ of $A \in \operatorname{Mat}(n, \mathbb{K})$ is the set of its complex eigenvalues, For an endomorphism $\varphi : E \to E$ of a finite-dimensional $\mathbb{K}$-vector space E, $\operatorname{spec}(\varphi)$ is defined in the same way. Here $E = T_m M$ or $E = T_m S_\varepsilon$, respectively.

half of Figure 7.3.1, the occurrence of an eigenvalue -1 could give rise to an orbit of twice the period.

In Chapters 7 and 8 of ABRAHAM and MARSDEN [AM], one finds illustrations for generic and Hamiltonian bifurcations respectively. ◊

7.3.3 Bifurcations of the Phase Space

As mentioned in the beginning of this chapter, the situation could occur in Hamiltonian systems with Hamilton function $H : M \to \mathbb{R}$ that the (reduced) phase space $H^{-1}(E)$ itself 'changes its form' in dependence on a parameter, since the energy, being a constant of motion, can be viewed as a parameter.

Dependent on the structure of the dynamics, further constants of motion $F_k : M \to \mathbb{R}$ can exist, and we will combine them with H to a *single* mapping $F \in C^\infty(M, N)$.

Whenever such a mapping F of differentiable manifolds is given, the question arises how the level sets $F^{-1}(f)$ change with $f \in N$.

In the simplest, but untypical, case, $F : M \to N$ is a C^∞-fiber bundle in the sense of Definition F.1 in the appendix. Then N can be viewed as the base of the bundle, and all the fibers $F^{-1}(f)$ are diffeomorphic to one standard fiber.

More frequently, this situation occurs only locally (see also [AM], *Section* 4.5):

7.23 Definition *For manifolds M and N, let $F \in C^\infty(M, N)$.*

- *F is called* **locally trivial at** *$f_0 \in N$, if there is a neighborhood $V \subseteq N$ of f_0 such that:*
  - *For all $f \in V$, the set $F^{-1}(f) \subset M$ is a differentiable submanifold;*
  - *There is a mapping $G \in C^\infty\big(U, F^{-1}(f_0)\big)$ on $U := F^{-1}(V)$, for which $F \times G : U \to V \times F^{-1}(f_0)$ is a diffeomorphism.*
- *The* **bifurcation set** *of F is*

$$\mathcal{V}(F) := \{ f \in N \mid F \text{ is not locally trivial at } f \} .$$

Note first that at a locally trivial point, the restricted mapping $F\restriction_U : U \to V$ is a C^∞-fiber bundle, even a product bundle.

On the other hand, for singular values $f \in N$ of F (i.e., those for which there is a pre-image $m \in M$ at which the linear mapping $T_m F : T_m M \to T_n N$ is not regular, see Definition A.45):

7.24 Lemma *The singular values of F belong to the bifurcation set $\mathcal{V}(F)$.*

Proof: Let $m_0 \in M$ and $f_0 := F(m_0)$. We assume that F is locally trivial at $f_0 \in N$. Then $\operatorname{rank}\big(T_{m_0}(F, G)\big) = \dim(M)$. Since

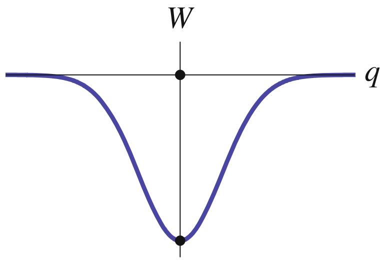
$$\operatorname{rank}\big(T_{m_0}(G)\big) \le \dim\big(F^{-1}(f_0)\big) = \dim(M) - \dim(N) ,$$

$\text{rank}\big(T_{m_0}(F)\big) = \dim(N)$, and thus m_0 is a regular point. □

One might conjecture that $\mathcal{V}(F)$ actually consists *only* of the singular values. But the following example shows that this is not necessarily the case:

7.25 Example (Bifurcation Set)
For $W : \mathbb{R} \to \mathbb{R}$, $q \mapsto -\exp(-q^2)$, (see top figure) only the minimal value -1 is singular, because $W'(q) = -2q\, W(q)$ vanishes only when $q = 0$.

The bifurcation set of W is however $\mathcal{V}(W) = \{-1, 0\}$, because $W^{-1}(E) = \emptyset$ for $E \geq 0$ and $|W^{-1}(E)| = 2$ for $E \in (-1, 0)$. To see that the values $E \in (-1, 0)$ are locally trivial, we use, in Definition 7.23, the sets $V := (-1, 0)$ and $U = F^{-1}(V) = \mathbb{R}^*$, and the mapping

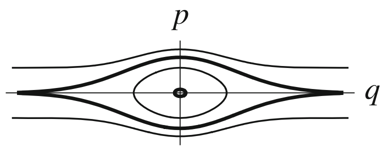


$$G : \mathbb{R}^* \to W^{-1}(E)\,,\qquad q \mapsto \text{sign}(q)\sqrt{\ln(1/|E|)}\,.$$

If we understand W as a potential of the Hamilton function

$$H : \mathbb{R}_p \times \mathbb{R}_q \to \mathbb{R},\ H(p, q) = \tfrac{1}{2}p^2 + W(q),$$

then $\mathcal{V}(H) = \{-1, 0\}$ as well.



Level sets of H are displayed in the lower figure. Their topological type changes at the bifurcation set: Whereas the level sets for $E \geq 0$ are diffeomorphic to two copies of $\mathbb{R}$, $H^{-1}(E)$ is diffeomorphic to a circle S^1 for $E \in (-1, 0)$. For $E < -1$ in turn, $H^{-1}(E) = \emptyset$. ◊

In Section 11.3.2, we will study bifurcations for the (integrable) two center model of celestial mechanics.

7.26 Literature A more extensive look at bifurcation theory is found in GUCKENHEIMER and HOLMES [GuHo]. The monograph [MMC] by MARSDEN and MCCRACKEN on the Hopf bifurcation can also be obtained online. MARX and VOGT [MV] combine theory and numerics of bifurcations. ◊

Chapter 8
Variational Principles

Parabola slides, Zentrum Mathematik, at the TU München
(Technical University of Munich, Germany).[1]

The Lagrange equations arising from a Lagrange function are second order differential equations. With this formalism, it is possible to realize constraints (such as occur in applications when objects are affixed to an axle or connected by rods) by simply restricting the Lagrange function.

[1]Image: courtesy of Zentrum Mathematik (Technical University of Munich, Germany).

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_8

Variational principles interpret solutions to differential equations as extrema of functions that are defined on spaces of curves. This interpretation simplifies the task of finding solutions, but also is a conceptual preparation for quantum mechanics.

8.1 Lagrange and Hamilton Equations

So far, we have adopted the point of view that the goal of analytical mechanics is to find the solutions to the Hamiltonian differential equation, given the Hamiltonian, or, should this turn out to be impossible, at least to study qualitative properties like fixed points, stability, etc.

It has not been analyzed how to find a Hamiltonian function that describes a given mechanical system.

Certainly, to some extent, this question is not a mathematical one, but rather belongs to the domain of modeling in physics. On the other hand, nevertheless, it is possible to calculate the Hamiltonian of a system in physics, if only a certain form of elementary interactions (like electromagnetism, or gravitation) is assumed.

In doing so, however, there is a problem related to the notion of momentum. Whereas position and velocity of a particle are well-defined quantities that can be measured,[2] the momentum is a dependent quantity for which a measuring protocol can only be found *after* the Hamiltonian H is known. The velocity is $\dot{q} = \mathrm{D}_1 H(p, q)$. Therefore the momentum is proportional to the velocity only if the Hamiltonian is of the form $H(p, q) = c\|p\|^2 + V(q)$.

This is one of the reasons why, next to the Hamilton function, also the Lagrange function is important in mechanics. The latter is a function of positions and velocities (rather than positions and momenta).

In the example just considered, the associated Lagrange function is

$$L(q, \dot{q}) = \tfrac{1}{4c}\|\dot{q}\|^2 - V(q)\,. \tag{8.1.1}$$

The first term in the sum is also called *kinetic energy*, the second *potential energy* of the particle. As both terms are subtracted, rather than added, the Lagrange function is in general not a constant of motion. This disadvantage is contrasted however (as will be shown in Section 8.2) by the fact that constraints can be incorporated easily into the Lagrange formalism.

In any case, one can typically switch between these two formalisms of mechanics (see Theorem 8.6).

[2] This also applies in special relativity, in each inertial frame, and one can convert between different inertial frames.

The most direct, and historically first, formalism of mechanics is the one by Newton, namely the law

$$\text{Force} = \text{Mass} \times \text{Acceleration}\,, \quad \text{or} \quad F(q,\dot q,t) \;=\; m\,\ddot q\,.$$

This formalism gives the equations of motion *immediately*.

As seen in the introduction already, it is useful to observe that the force frequently is not just a function of the position q, but also that it can be represented as the (negative) gradient of a real-valued function V of q:

$$F(q) = -\nabla V(q)\,. \tag{8.1.2}$$

From this fact, and with the momentum $p = mv$, we conclude that Newton's equations arise as Hamilton equations from the Hamilton function $H(p,q) = \frac{\|p\|^2}{2m} + V(q)$. From Theorem 6.3, one concludes:

8.1 Corollary *The force field $F \in C^1(U,\mathbb{R}^n)$ on an open and simply connected (e.g., convex) configuration space $U \subseteq \mathbb{R}^n_q$ is a gradient vector field if and only if its components satisfy*

$$\frac{\partial F_i}{\partial q_k} = \frac{\partial F_k}{\partial q_i} \qquad (i,k \in \{1,\ldots,n\}).$$

For instance, when $n = 3$, this means that the curl of F has to vanish.

So it is a specific form of the forces that allows to express the mechanical equations of motion in terms of a single function.

8.2 Definition

- *For a function $L \in C^2\left(U \times \mathbb{R}^n_v, \mathbb{R}\right)$, called* **Lagrange function**, *or* **Lagrangian**, *with an open configuration space $U \subseteq \mathbb{R}^n_q$, the system of differential equations*

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial v_i}(q,\dot q) = \frac{\partial L}{\partial q_i}(q,\dot q) \qquad (i = 1,\ldots,n)} \tag{8.1.3}$$

  is called **Lagrange equation(s)** *of L.*
- $p := \mathrm{D}_2 L(q,v)$ *is called* **(generalized) momentum**.

8.3 Example The choice $c := 1/(2m)$ in (8.1.1) yields $L(q,v) = \frac{m}{2}\|v\|^2 - V(q)$. So the Lagrange equation $m\ddot q = -\nabla V(q)$ corresponds to the Newtonian equation of motion of a particle with mass m in the potential V. ◊

In this example, we can solve the equation $p = \mathrm{D}_2 L(q,v)$ defining the momentum for v, namely $v \equiv v(p,q) = p/m$. If we now consider the Hamiltonian

$$H(p,q) := \langle p, v(p,q)\rangle - L\big(q, v(p,q)\big) = \frac{\|p\|^2}{2m} + V(q)\,, \tag{8.1.4}$$

and rewrite the Lagrange equations as a system of $2n$ first order equations, using the momentum, one obtains that these equations

$$\left.\begin{aligned} m\dot{v} &= -\nabla V(q) \\ \dot{q} &= v \end{aligned}\right\} \iff \left\{\begin{aligned} \dot{p} &= -\nabla V(q) \\ \dot{q} &= p/m \end{aligned}\right.$$

are the Hamiltonian equations of H.

We want to make such a connection for other Lagrangians as well. The question arises which hypotheses on the Lagrangian L allow for the equation $p = p(q, v) = D_2 L(q, v)$ to be rewritten in such a way that $v = v(p, q)$ becomes the dependent variable.

This geometric problem of a variable transformation defined in terms of derivatives is solved by the Legendre transform (see Appendix C).

For those Lagrangians that satisfy the relevant condition, (8.1.4) will yield a Hamilton function whose Hamiltonian differential equation is equivalent to the Lagrange equation of L.

8.4 Example (Legendre Transform of Quadratic Forms)
For a symmetric matrix $A \in \mathrm{M}at(d, \mathbb{R})$, $A > 0$, the quadratic form

$$f : \mathbb{R}^d \to \mathbb{R} \quad , \quad f(x) := \tfrac{1}{2} \langle x, Ax \rangle$$

has the second derivative $D^2 f(x) \equiv A$ and the Legendre transform

$$f^*(p) = \sup_{x \in \mathbb{R}^d} \big(\langle p, x \rangle - f(x)\big) = \langle p, x(p) \rangle - \tfrac{1}{2} \langle x(p), Ax(p) \rangle \quad \text{with } x(p) = A^{-1} p,$$

hence $f^* : \mathbb{R}^d \to \mathbb{R}$, $f^*(p) = \frac{1}{2} \langle p, A^{-1} p \rangle$. This relation serves, e.g., to convert the kinetic energy from velocity coordinates to momentum coordinates. ♢

8.5 Exercise (Legendre Transform) Let $H : \mathbb{R}^d \to \mathbb{R}$ be convex and

$$H^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\} \quad , \quad H^*(q) := \sup_{p \in \mathbb{R}^d} \big(\langle p, q \rangle - H(p)\big)$$

the Legendre transform of H.

(a) Show: If $H(p) = \frac{1}{r} \|p\|^r$ with $r \in (1, \infty)$, then $H^*(q) = \frac{1}{s} \|q\|^s$, with $\frac{1}{r} + \frac{1}{s} = 1$.

(b) Now choose $H(p) = \frac{1}{2} \langle p, Ap \rangle + \langle b, p \rangle + c$ with a symmetric, positive definite matrix A, $b \in \mathbb{R}^d$ and $c \in \mathbb{R}$. Determine H^*.
Remark: This formula is used for the Legendre transform of the Lagrangian that describes the motion in an electromagnetic field, see Exercise 8.8 (b). ♢

More generally, we now apply the Legendre transform to convert the Lagrangian into the Hamiltonian, and conversely. This can be done if the growth in the velocity is quadratic[3]:

8.6 Theorem
For $L \in C^2(U \times \mathbb{R}^n_v, \mathbb{R})$ with an open $U \subseteq \mathbb{R}^n_q$, let there exist an $a > 0$ such that

$$\big\langle w, \mathrm{D}_v^2 L(q,v), w\big\rangle \geq a\,\langle w, w\rangle \qquad \big((q,v) \in U \times \mathbb{R}^n_v,\ w \in \mathbb{R}^n\big).$$

• *Then, with the momentum $p \equiv p(q,v) = \mathrm{D}_v L(q,v)$, the function*

$$\boxed{H(p,q) = \langle p, v\rangle - L(q,v)}$$

is the Legendre transform of L with respect to v, and $H \in C^2(\mathbb{R}^n_p \times U, \mathbb{R})$.
• *The Lagrange equations (8.1.3) are equivalent to the Hamilton equations*

$$\dot p_i = -\frac{\partial H}{\partial q_i} \quad , \quad \dot q_i = \frac{\partial H}{\partial p_i} \qquad (i = 1, \ldots, n).$$

Proof: • In the relation between H and L, the quantity $q \in U$ plays merely the role of a parameter. Letting $\mathcal{L}_q(v) := L(q,v)$, Theorem C.7 tells us that $v \mapsto p(q,v) = \mathrm{D}\mathcal{L}_q(v)$ is a diffeomorphism onto the image. This image $B := p(q, \mathbb{R}^n) \subseteq \mathbb{R}^n$ is however again $\mathbb{R}^n$ due to the hypothesis $\mathrm{D}_v^2 L(q,v) \geq a\,1\!\!1$.
• Let us consider the mapping

$$\Phi : U \times \mathbb{R}^n_v \to \mathbb{R}^n_p \times U \quad , \quad \left(\begin{smallmatrix} q \\ v \end{smallmatrix}\right) \mapsto \left(\begin{smallmatrix} p \\ q \end{smallmatrix}\right) = \left(\begin{smallmatrix} \mathrm{D}_v L(q,v) \\ q \end{smallmatrix}\right).$$

By what has just been proved about the mapping $v \mapsto p(q,v)$, the mapping Φ is once continuously differentiable and bijective. At the point (q,v) with image (p,q), its total derivative is given as

$$\mathrm{D}\Phi(q,v) = \begin{pmatrix} \mathrm{D}_q \mathrm{D}_v L(q,v) & \mathrm{D}_v^2 L(q,v) \\ 1\!\!1 & 0 \end{pmatrix}.$$

The Jacobi matrix has the determinant $(-1)^n \det\big(\mathrm{D}_v^2 L\big) \neq 0$ and is therefore invertible. Then, by Theorem 2.38, Φ is a locally invertible, hence a local diffeomorphism; and being bijective, it is a diffeomorphism.
• It follows from Theorem C.9 that H has the same order of differentiability as L does.
• The equivalence of the equations of motion follows by taking the total exterior derivatives of H and L: On one hand, $dH = \mathrm{D}_p H\, dp + \mathrm{D}_q H\, dq$, on the other hand in view of $\mathrm{D}_v L = p$, the exterior derivative $dH = d\big(\langle v, p\rangle - L(q,v)\big)$ equals

[3] As the statement of Exercise 8.5 (a) indicates, this condition is by no means necessary.

$$v\,dp + p\,dv - \mathrm{D}_q L\,dq - \mathrm{D}_v L\,dv = v(q,p)\,dp - \mathrm{D}_q L\,dq\,.$$

Comparing coefficients yields $\dot{q} = v = \mathrm{D}_p H$, $\mathrm{D}_q H = -\mathrm{D}_q L = -\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}_v L = -\dot{p}$, if we assume the Lagrange equations $\mathrm{D}_q L(q,v) = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}_v L(q,v)$ (along with the definition $v = \frac{\mathrm{d}q}{\mathrm{d}t}$ of the velocity). □

Conversely, from a Hamiltonian H that is convex in the momentum p, we can obtain the Lagrangian by means of the Legendre transform with respect to p.

8.7 Example (Relativistic Hamiltonian)
If $c > 0$ denotes the speed of light, the Hamiltonian of a free relativistic particle with mass $m > 0$ is

$$H : \mathbb{R}^3_p \times \mathbb{R}^3_q \to \mathbb{R} \quad , \quad H(p,q) = c\sqrt{\|p\|^2 + m^2c^2}\,.$$

Interpreting H as the total energy, we obtain a Taylor expansion

$$H(p,q) = mc^2 + \frac{\|p\|^2}{2m} + \mathcal{O}(\|p\|^4)\,,$$

in which the first term is called *rest energy* and the second term is the non-relativistic kinetic energy. Here the velocity

$$\dot{q} = \mathrm{D}_p H(p,q) = c\frac{p}{\sqrt{\|p\|^2 + m^2c^2}}$$

can take on only values of absolute value strictly less than the speed of light c. The Lagrangian $L : \mathbb{R}^n_q \times U_c(0) \to \mathbb{R}$ equals

$$L(q,v) = \langle v, p(v)\rangle - H\big(p(v),q\big) = -\frac{m^2c^3}{\sqrt{\|p(v)\|^2 + m^2c^2}} = -mc^2\sqrt{1 - \frac{\|v\|^2}{c^2}}.$$

In this example, the domain of L is not all of $\mathbb{R}^n_q \times \mathbb{R}^n_v$, as $\mathrm{D}^2_p H$ fails to satisfy the hypothesis of Theorem 8.6. ◊

8.8 Exercise (Legendre-Transformation)

(a) A particle is moving in the configuration space $\mathbb{R}^2$ in a central potential

$$V : \mathbb{R}^2 \to \mathbb{R} \quad , \quad V(q) = U(\|q\|) \quad \text{with} \quad U : \mathbb{R} \to \mathbb{R}\,.$$

Re-express the Lagrangian $\tilde{L}(q,v) = \frac{1}{2}\|v\|^2 - V(q)$ in terms of planar polar coordinates (r,φ). Let the result be called L. Calculate the Legendre transform of L with respect to velocity (v_r, v_φ).

(b) The motion of a charged particle in an electromagnetic field is described by the Lagrangian

$$L \in C^\infty(\mathbb{R}^n_q \times \mathbb{R}^n_v, \mathbb{R}) \quad , \quad L(q,v) = \tfrac{1}{2} m \|v\|^2 - e\phi(q) + \tfrac{e}{c} \langle v, A(q) \rangle \ ,$$

where $\phi \in C^\infty(\mathbb{R}^n, \mathbb{R})$ and $A \in C^\infty(\mathbb{R}^n, \mathbb{R}^n)$. Here $e \in \mathbb{R}$ is the charge of the particle and c the speed of light. Calculate the Hamiltonian by means of a Legendre transform. ♢

8.2 Holonomic Constraints

In many cases, the motion of the particles to be described is restricted in some way or another. For instance, we will discuss the example of a bead on a wire.

Let us consider the motion of a particle in an m-dimensional submanifold $S \subseteq \mathbb{R}^n_q$. Let the Lagrangian of the particle in $\mathbb{R}^n_q \times \mathbb{R}^n_v$ be $\tilde{L}$. Near a point $q_0 \in S$, we can parametrize S locally as $q = q(x)$, $x = (x_1, \ldots, x_m)$.

8.9 Definition *With $L(x,w) := \tilde{L}\big(q(x), \mathrm{D}q(x)\, w\big)$, the ODE of second order*

$$\frac{\mathrm{d}}{\mathrm{d}t} D_2 L(x, \dot{x}) = D_1 L(x, \dot{x})$$

is called system with **configuration space** *S and* **holonomic constraint**.

8.10 Remark (**Non-holonomic Constraints**)
By restricting the location q to the submanifold $S \subseteq \mathbb{R}^n_q$, the velocity is restricted to the subspace $T_q S$ of the tangential space $T_q \mathbb{R}^n_q$.

In a neighborhood $U \subseteq \mathbb{R}^n_q$ of Q, the manifold S can be written as the level set $S = F^{-1}(0)$ of the regular value 0 of some function $F \in C^\infty(U, \mathbb{R}^{n-m})$.
This function defines the geometric distribution $\mathcal{D}_F$ on U (see Definition F.23):

$$\mathcal{D}_F := \big\{(u,v) \mid u \in U,\ v \in T_u U,\ (\mathrm{D}F)_u(v) = 0\big\} \subseteq TU \ .$$

This is a smooth family of subspaces $\ker(\mathrm{D}F)_u \subset T_u U$. So if we consider, for an initial configuration $q_0 \in U$ not necessarily lying on S, only those trajectories $c : I \to U$ with $c(0) = q_0$ and $\big(c(t), c'(t)\big) \in \mathcal{D}_F$, then these trajectories will accordingly remain on the level set $F^{-1}(f)$ of $f := F(q_0)$.
Such a constraint that is defined by the distribution $\mathcal{D}_F$ is also called *holonomic*.

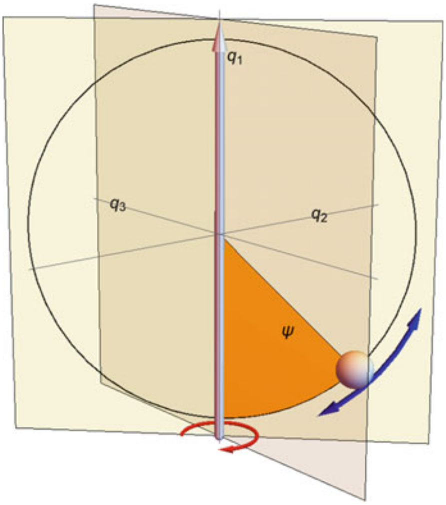
We are here not interested in the question which choices of f can be realized in the context of physics. An example would be $F : \mathbb{R}^2 \to \mathbb{R}$, $F(q) = \|q\|^2$, i.e., a pendulum in the plane with adjustable length f.

However, what is of interest is the following generalization:
Define a *constraint* as a distribution $\mathcal{D} \subseteq TU$ in the configuration space. We call the constraint holonomic if this distribution is integrable, and *non-holonomic* otherwise. Non-holonomic constraints occur in the example of a rolling ball (see Chapter 14.4). ♢

If we want to describe the motion of a particle on a submanifold S of the configuration space, we first give the Lagrangian $\tilde{L}$ of the particle that can move freely in $\mathbb{R}^n_q$, then calculate L, and from it, if desired, the Hamiltonian $H : T^*S \to \mathbb{R}$ by means of the Legendre transform. Here T^*S is called the cotangent bundle of S, which is a $2m$-dimensional manifold whose points are pairs (p, x) with $x \in S$ and p the canonical momentum.

While initially the above construction is only valid in a neighborhood of $q_0 \in S$, we will discuss the notion of a manifold in Appendix A.2, and it will allow us to discuss the global motion on S.

8.11 Example (**Bead on a Wire**)
A bead slides on a circular wire without friction. The circle of radius R is centered at $0 \in \mathbb{R}^3$. It rotates with angular velocity ω about the (vertical) q_1-axis, and the plane in which the circle lies contains this axis (see figure).



Denoting by ψ the angle between the lower equilibrium position and the actual position q of the bead, one has

$$q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = R \begin{pmatrix} -\cos(\psi) \\ \sin(\psi)\,\sin(\omega t) \\ \sin(\psi)\,\cos(\omega t) \end{pmatrix}.$$

The Lagrangian of the bead is thus of the form $\tilde{L}(q, v) := \frac{m}{2}\|v\|^2 - V(q)$, namely it is the difference between kinetic and potential energy. The latter is equal to $V(q) = mgq_1$, where $g > 0$ is the acceleration of gravity as usual.

The bead has one, rather than three, degrees of freedom. Accordingly, we write the Lagrangian as a function of ψ, the angular velocity $v_\psi = \dot{\psi}$, and possibly the time t (see also Percival and Richards [PR2]). We get

$$\dot{q} = \begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{pmatrix} = R \begin{pmatrix} \sin(\psi)\dot{\psi} \\ \cos(\psi)\sin(\omega t)\dot{\psi} + \omega\ \sin(\psi)\cos(\omega t) \\ \cos(\psi)\cos(\omega t)\dot{\psi} - \omega\ \sin(\psi)\sin(\omega t) \end{pmatrix},$$

hence

$$\|\dot{q}\|^2 = R^2\left(\dot{\psi}^2 + \omega^2 \sin^2(\psi)\right).$$

Moreover, $V(q) = -mgR\cos(\psi)$, and therefore in the new coordinates, the Lagrangian takes the (time-independent) form

$$L(\psi, v_\psi) = \frac{mR^2}{2} v_\psi^2 + mR\left(g\cos(\psi) + \frac{\omega^2 R}{2}\sin^2(\psi)\right).$$

The conjugate momentum p_ψ for the angle ψ is therefore

$$p_\psi := \frac{\partial L}{\partial v_\psi} = mR^2\, v_\psi\,.$$

Thus the Hamiltonian has the form

$$H(p_\psi,\psi) \;=\; v_\psi\, p_\psi - L(\psi, v_\psi) \;=\; \frac{p_\psi^2}{2mR^2} - mR\left(g\cos(\psi) + \tfrac{1}{2}\omega^2 R\sin^2(\psi)\right).$$

We may restrict our discussion to the case $m = R = g = 1$, which can always be achieved by appropriately rescaling length, energy, and time. We then get, with a new angular velocity $\Omega := \omega\,\sqrt{R/g}$, the new Hamiltonian

$$H_\Omega(p_\psi,\psi) = \tfrac{1}{2}p_\psi^2 + W_\Omega(\psi) \qquad \big((p_\psi,\psi)\in M\big)$$

with the potential $W_\Omega(\psi) := -\cos(\psi) - \frac{\Omega^2}{2}\sin^2(\psi)$. The phase space of H_Ω is the cylinder $M := \mathbb{R}\times S^1$.

For $\Omega = 0$, one has $W_0(\psi) = -\cos(\psi)$, so the potential has a non-degenerate minimum in the location of the lower equilibrium $\psi = 0$ (Figure 8.2.1).

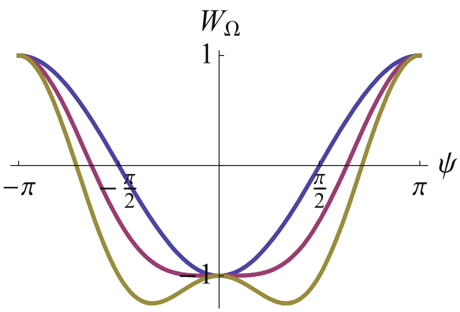


Figure 8.2.1 Potential W_Ω for $\Omega = 0$, $\Omega = 1$ and $\Omega = \sqrt{2}$

On the other hand, for larger $|\Omega|$, the centrifugal force becomes dominant, and $W_\Omega(\psi)$ has minima for values of ψ that are close to $\pm\pi/2$ (see Figure 8.2.1).

In the phase space portrait of H_Ω, this translates into a bifurcation (see Figure 8.2.2). For all values of Ω, the lower equilibrium in M, namely the point with the coordinates $(p_\psi,\psi) = (0,0)$, is a zero of the vector field

$$X_{H_\Omega}(p_\psi,\psi) = \begin{pmatrix} -W_\Omega'(\psi) \\ p_\psi \end{pmatrix} = \begin{pmatrix} -\sin(\psi) + \frac{1}{2}\Omega^2\sin(2\psi) \\ p_\psi \end{pmatrix} \qquad \big((p_\psi,\psi)\in M\big)$$

associated with H_Ω. We can calculate its stability by studying the linearization $DX_{H_\Omega}(p_\psi,\psi) = \begin{pmatrix} 0 & -W_\Omega''(\psi) \\ 1 & 0 \end{pmatrix}$. The eigenvalues of DX_{H_Ω} are $\pm\sqrt{-\det(DX_{H_\Omega})}$.

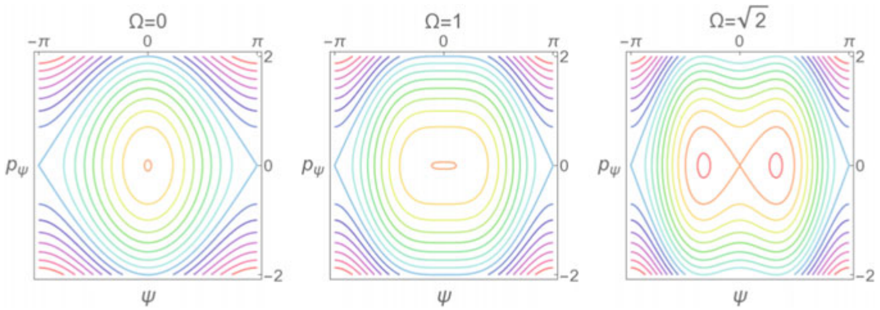


Figure 8.2.2 Phase space portraits for H_Ω

Because of

$$\det\big(DX_{H_\Omega}\big)(\psi) = W''_\Omega(\psi) = \cos(\psi) - \Omega^2\cos(2\psi)\,,$$

they are equal to $\pm\sqrt{-1+\Omega^2}$, namely imaginary for $|\Omega| \le 1$ and real for $|\Omega| \ge 1$. Thus for angular velocities whose absolute value is below the bifurcation value 1, the lower equilibrium will be Lyapunov-stable according to Remark 7.9.1, whereas for larger values of the angular velocity, it will be unstable. ◊

8.12 Exercise (**Bead on a Wire**) A bead with mass m slides without friction under the influence of acceleration of gravity g on a parabolic wire of the form $z = \frac{1}{2}\alpha^2 x^2$, where the z-axis points vertically upwards. The wire rotates with a constant angular velocity ω about the z-axis.

(a) Calculate the Hamiltonian H.
(b) Investigate the stability of the point $(0, 0)$ by means of the linearization of the Hamiltonian vector field X_H.
(c) Show that for $g\alpha^2 > \omega^2$ (i.e., slow rotation) and energy E, the bead performs a periodic motion with period

$$T = \frac{\sqrt{8m}}{\beta}\int_0^{\frac{\pi}{2}} \left(1 + \frac{\alpha^4 E}{\beta^2}\sin^2(\theta)\right)^{1/2} d\theta$$

with $\beta^2 = \frac{m}{2}(g\alpha^2 - \omega^2)$. ◊

8.3 The Hamiltonian Variational Principle

The shortest connection between two points $x_0, x_1 \in \mathbb{R}^n$ is the segment connecting these points. This can be proved by considering all possible paths between the two points. If these paths γ are continuously differentiable with respect to their parameter, they have a well-defined length:

$$\gamma : [0, 1] \to \mathbb{R}^n \quad , \quad \gamma(0) = x_0,\ \gamma(1) = x_1 \quad , \quad I(\gamma) := \int_0^1 L\big(\dot{\gamma}(t)\big)\,\mathrm{d}t$$

with Lagrangian

$$L(v) := \|v\| = \sqrt{v_1^2 + \ldots + v_n^2}\,.$$

The *length functional* I is therefore a mapping from the set of eligible paths into the real numbers. Such a functional, just like a function defined on a finite dimensional space, can have minima, or, more generally, extrema. And just as in the finite dimensional case, the linearization at the extremum will be the zero mapping: The *variation* of I vanishes there.

Let us generalize this approach. In doing so, we will assume that functions considered in the sequel are smooth enough for our purposes. L will be a general Lagrangian that may depend on time.

So if configuration space is an open set $U \subseteq \mathbb{R}^n_q$, then we consider a Lagrange function

$$L \in C^2\big(U \times \mathbb{R}^n_v \times [t_0, t_1], \mathbb{R}\big)\,.$$

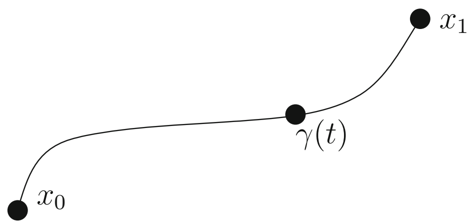
Choosing (for instance) the space

$$X := \big\{\gamma \in C^2\big([t_0, t_1], U\big) \mid \gamma(t_0) = x_0, \gamma(t_1) = x_1\big\}$$

of paths from x_0 to $x_1 \in U$, we use the action functional

$$I : X \to \mathbb{R} \quad , \quad I(\gamma) = \int_{t_0}^{t_1} L\big(\gamma(t), \dot{\gamma}(t), t\big)\,\mathrm{d}t\ .$$

The difference $h := \gamma_1 - \gamma_2$ of two paths with common start point x_0 and end point x_1 is a mapping h belonging to an $\mathbb{R}$-vector space, namely to



$$X_0 := \big\{h \in C^2\big([t_0, t_1], \mathbb{R}^n\big) \mid h(t_0) = h(t_1) = 0\big\}\,.$$

Here X_0 is neither finite dimensional (as is seen easily in a Fourier representation of $h \in X_0$), nor is X a linear space.

On the vector space X_0, we can, for instance, introduce the norm

$$\|h\|_0 := \sup_{t\in[t_0,t_1]} \big(\|h(t)\| + \|\mathrm{D}h(t)\|\big)\,.$$

This norm leads to the metric $d(\gamma_1, \gamma_2) := \|\gamma_1 - \gamma_2\|_0$ on X. The image of a curve $\gamma \in X$ is a compact subset of $U \subseteq \mathbb{R}^n$, and the complement $\mathbb{R}^n \setminus U$ is closed. So there exists an $\varepsilon > 0$ such that the image of $(\gamma + h) : [t_0, t_1] \to \mathbb{R}^n$ lies in U if $\|h\|_0 < \varepsilon$.

8.13 Definition *A functional $I : X \to \mathbb{R}$ is called*

- **(Fréchet) differentiable at** $\gamma \in X$ *if there exists a bounded linear mapping $\mathcal{L}_\gamma : X_0 \to \mathbb{R}$ such that*

$$I(\gamma + h) - I(\gamma) = \mathcal{L}_\gamma(h) + o(\|h\|_0) \qquad (h \in X_0 \text{ with } \gamma + h \in X). \tag{8.3.1}$$

- *I is called* **(Fréchet) differentiable** *if $\mathcal{L}_\gamma$ exists for all $\gamma \in X$.*
- *γ is called a* **stationary point** *(or* **critical point***) of I if $\mathcal{L}_\gamma = 0$.*
- *I is called* **continuously differentiable** *if even the mapping $\gamma \mapsto \mathcal{L}_\gamma$ is continuous.*

8.14 Notation (**Functional Derivative**) If the linear mapping $\mathcal{L}_\gamma$ exists, then it is uniquely defined by (8.3.1). More common than $\mathcal{L}_\gamma$ are the notations $\delta I(\gamma)$ and $\mathrm{D}I(\gamma)$.[4] With this notation, $\mathrm{D}I(\gamma) = 0$ for stationary points γ.

8.15 Theorem *For $L \in C^2(U \times \mathbb{R}^n_v \times \mathbb{R}_t, \mathbb{R})$ and $\tilde{\gamma}(t) := \big(\gamma(t), \dot{\gamma}(t), t\big)$,*

$$I : X \to \mathbb{R} \quad , \quad I(\gamma) := \int_{t_0}^{t_1} L\big(\tilde{\gamma}(t)\big)\,\mathrm{d}t \tag{8.3.2}$$

is a differentiable functional, whose derivative is

$$(\mathrm{D}I)_\gamma(h) = \int_{t_0}^{t_1} \Big[\mathrm{D}_1 L - \tfrac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}_2 L\Big] \circ \tilde{\gamma}(t)\; h(t)\,\mathrm{d}t \qquad (h \in X_0). \tag{8.3.3}$$

Proof:

• The left side of the equation (8.3.1) defining $\mathcal{L}_\gamma$ is

$$\begin{aligned} I(\gamma + h) - I(\gamma) &= \int_{t_0}^{t_1} L\big(\gamma(t) + h(t), \dot{\gamma}(t) + \dot{h}(t), t\big)\,\mathrm{d}t - \int_{t_0}^{t_1} L\big(\tilde{\gamma}(t)\big)\,\mathrm{d}t \\ &= \int_{t_0}^{t_1} \big[\mathrm{D}_1 L(\tilde{\gamma}(t)) \cdot h(t) + \mathrm{D}_2 L(\tilde{\gamma}(t)) \cdot \dot{h}(t)\big]\,\mathrm{d}t + \mathcal{O}\left(\|h\|_0^2\right). \end{aligned}$$

By integration by parts, we get rid of the time derivative of the variation h:

$$\int_{t_0}^{t_1} \mathrm{D}_2 L\big(\tilde{\gamma}(t)\big)\, \dot{h}(t)\,\mathrm{d}t = \mathrm{D}_2 L \cdot h\big|_{t_0}^{t_1} - \int_{t_0}^{t_1} \tfrac{\mathrm{d}}{\mathrm{d}t}\big(\mathrm{D}_2 L(\tilde{\gamma}(t))\big) \cdot h(t)\,\mathrm{d}t\,,$$

Since $h \in X_0$, we have $h(t_0) = h(t_1) = 0$, hence $(\mathrm{D}I)_\gamma(h)$ is indeed given by (8.3.3).

• $\mathrm{D}I(\gamma) : X_0 \to \mathbb{R}$ is bounded, with operator norm

$$\|\mathrm{D}I(\gamma)\| = \sup_{h \in X_0,\, \|h\|_0 = 1} |(\mathrm{D}I)_\gamma(h)| \le (t_1 - t_0) \sup_{t \in [t_0, t_1]} \left\| \big(\mathrm{D}_1 L - \tfrac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}_2 L\big) \circ \tilde{\gamma}(t) \right\|,$$

because the vector valued function is continuous on the compact set $[t_0, t_1]$. □

---

[4]When we apply the operator $\mathrm{D}I(\gamma)$ to a vector h, we write γ as an index.

8.16 Theorem (**Hamiltonian Variational Principle**)
$\gamma \in X$ is a stationary point of the functional (8.3.2) if and only if γ satisfies the **Euler-Lagrange Equation**

$$\left(\mathrm{D}_1 L - \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}_2 L\right) \circ \tilde{\gamma}(t) = 0 \qquad \big(t \in [t_0, t_1]\big). \tag{8.3.4}$$

To prove this theorem, we prove a simple lemma.

8.17 Lemma (**Fundamental Lemma of the Calculus of Variations**)
If $f \in C\big([t_0, t_1], \mathbb{R}\big)$ satisfies the condition $\int_{t_0}^{t_1} f(t)h(t)\,\mathrm{d}t = 0$ for all $h \in C\big([t_0, t_1], \mathbb{R}\big)$ with $h(t_0) = h(t_1) = 0$, then $f = 0$.

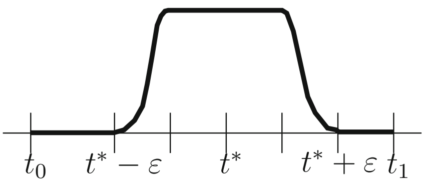
Proof: Assume that instead $f \neq 0$. Then, as f is continuous, there exists a t^* in the open interval (t_0, t_1) with $f(t^*) > 0$ (or < 0, with an analogous argument). Thus there exist $\varepsilon > 0$ and $c > 0$ such that $[t^* - \varepsilon, t^* + \varepsilon] \subset (t_0, t_1)$ and

$$f(t) \geq c \qquad \big(t \in [t^* - \varepsilon, t^* + \varepsilon]\big).$$

Choose a *cutoff function* $h \in C\big([t_0, t_1], \mathbb{R}\big)$ with $h \geq 0$,

$$h(t) = \begin{cases} 0\ , & t \leq t^* - \varepsilon \\ 1\ , & t^* - \frac{\varepsilon}{2} < t < t^* + \frac{\varepsilon}{2} \\ 0\ , & t \geq t^* + \varepsilon, \end{cases}$$

(see figure). Then $\int_{t_0}^{t_1} f(t)h(t)\,\mathrm{d}t \geq c\,\varepsilon > 0$. □



Proof of Theorem 8.16:
• If $\gamma \in X$ is a stationary point, i.e., $\mathrm{D}I(\gamma) = 0$, then it follows with (8.3.3) that

$$0 = (\mathrm{D}I)_\gamma(h) = \int_{t_0}^{t_1} \left[\mathrm{D}_1 L - \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{D}_2 L\right] \circ \tilde{\gamma}(t)\ h(t)\,\mathrm{d}t \quad (h \in X_0 \text{ with } \gamma + h \in X).$$

Applying Lemma 8.17 on the n components of $h \in X_0$ (which can be varied independently), yields the Euler-Lagrange equation (8.3.4).
• Conversely, the vanishing of $\mathrm{D}I(\gamma)$ follows from (8.3.4) according to (8.3.1). □

Thus the solutions of the Euler-Lagrange equation are precisely the extremals of the action functional I. This in turn is the integral of the Lagrangian along a path on phase space[5] $U \times \mathbb{R}^n \cong TU$, namely the tangential space of U. The argument of L is the curve $t \mapsto \tilde{\gamma}(t) = \big(\gamma(t), \dot{\gamma}(t)\big)$. Under a coordinate transformation of the configuration space with a smooth diffeomorphism $\psi : U \to V$, the curve $\gamma : [t_0, t_1] \to U$ becomes $\psi \circ \gamma : [t_0, t_1] \to V$, and $\tilde{\gamma}$ becomes

[5] We assume here for simplicity of the discussion that L does not explicitly depend on the time.

$$\widetilde{\psi \circ \gamma} : [t_0, t_1] \to V \times \mathbb{R}^n \quad , \quad \widetilde{\psi \circ \gamma}(t) = \big(\psi \circ \gamma(t),\ \mathrm{D}\psi_{\gamma(t)}\ \dot{\gamma}(t)\big) .$$

So it undergoes the transformation rule of a curve in the tangential space (see Theorem A.42), and we have proved:

8.18 Lemma *If γ satisfies the Euler-Lagrange equation for $L \in C^2\big(U \times \mathbb{R}^n_v, \mathbb{R}\big)$, then the same is true with $\psi \circ \gamma$ and $L \circ \Psi^{-1}$, where*

$$\Psi : U \times \mathbb{R}^n \to V \times \mathbb{R}^n \quad , \quad \Psi(q, v) := \big(\psi(q),\ \mathrm{D}\psi_q\ v\big) .$$

This behavior with respect to transformations allows us to also consider, for curves $\gamma \in C^2\big([t_0, t_1], M\big)$ in a configuration manifold M, the action functional

$$I(\gamma) := \int_{t_0}^{t_1} L\big(\gamma(t), \dot{\gamma}(t)\big)\, \mathrm{d}t$$

of Lagrange functions $L \in C^2(TM, \mathbb{R})$ on the tangent bundle TM of M, see Appendix A.3.

8.19 Remark (**Functionals**)

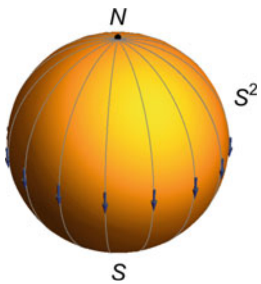
1. Not every functional has a stationary point (this is already true for functions, after all). As an example, take the length functional $L(x, v, t) := \|v\|$ in the punctured plane $U := \mathbb{R}^2 \setminus \{0\}$. The segment from $x_0 := \binom{-1}{0}$ to $x_1 := \binom{1}{0}$ contains the point $0 \in \mathbb{R}^2$. So within U, there is no shortest curve between x_0 and x_1.
2. Extremals need not be unique. As an example, consider the space

$$X := \left\{\gamma \in C^2\big([0, 1], S^2\big) \;\middle|\; \gamma(0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \ , \ \gamma(1) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}\right\}$$

of those curves that connect the north pole $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ to the south pole $\begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$ on the sphere $S^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$, and the functional

$$I(\gamma) := \tfrac{1}{2} \int_0^1 \left\| \frac{\mathrm{d}\gamma}{\mathrm{d}t}(t) \right\|^2 \mathrm{d}t .$$

The segments of great circles that begin at the north pole and end at the south pole are extremals of I (see adjacent figure, and also Example G.14.2).



Geodesics connecting north and south pole

3. Even if the image of a curve is unique, its parametrization need not be unique. An example is the length functional $I(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|\, \mathrm{d}t$, because under a diffeomorphism $c : [0, 1] \to [0, 1]$ with $c(0) = 0$ and $c(1) = 1$, it transforms as

$$\int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}t} \gamma\big(c(t)\big) \right\| \, \mathrm{d}t = \int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}s} \gamma(s) \right\| \, \mathrm{d}s \tag{8.3.5}$$

with the new parameter $s = c(t)$. Sometimes, this nonuniqueness of the parametrization causes trouble. This is why the *energy functional*

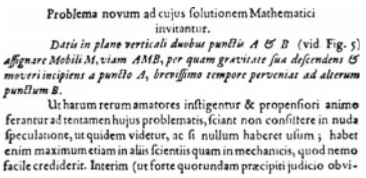
$$\gamma \mapsto \tfrac{1}{2} \int_0^1 \|\dot{\gamma}(t)\|^2 \, \mathrm{d}t \, ,$$

which has just been used in Remark 2, is preferred. Up to parametrization, it has the very same extremals as the length functional, but the parametrization is unique (see Remark G.3.2). ◊

8.20 Exercise (**Example on non-Minimality of the Action Functional**)
Given the Lagrange function

$$L \in C^\infty\big(\mathbb{R}^2 \times \mathbb{R}^2, \mathbb{R}\big) \quad , \quad L(q, v) := \tfrac{1}{2}\big(\|v\|^2 - q_2^2\big) \, ,$$

where $q = (q_1, q_2)^\top$. In physics, this could describe a particle moving in some gutter with a certain $\smile$ shaped cross section.


Problema novum ad cujus ſolutionem Mathematici invitantur.

Datis in plano verticali duobus punctis A & B (vid. Fig. 5) aſſignare Mobili M, viam AMB, per quam gravitate ſua deſcendens & moveri incipiens a puncto A, breviſſimo tempore perveniat ad alterum punctum B.

Ut harum rerum amatores inſtigentur & propenſiori animo ferantur ad tentamen hujus problematis, ſciant non conſiſtere in nuda ſpeculatione, ut quidem videtur, ac ſi nullum haberet uſum; habet enim maximum etiam in aliis ſcientiis quam in mechanicis, quod nemo facile crediderit. Interim (ut forte quorundam præcipiti judicio obvi-

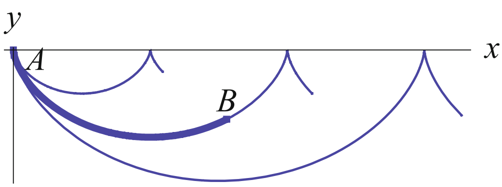


Figure 8.3.1 Left: Johann Bernoulli's announcement [Bern] about the brachystochrone problem. Right: A family of cycloids (to go with the box on page 169)

(a) Calculate the extremals q that satisfy $q(0) = (0, 0)^\top$ and $q(T) = (c, 0)^\top$ for the action functional $I(q) = \int_0^T L \, \mathrm{d}t$. (Be sure to take care to distinguish times $T = n\pi$ and $T \neq n\pi$.)

(b) Show, for the solution $q(t) = (\frac{c}{T}\, t, 0)^\top$ and any variation δq of q, i.e., for $\delta q \in C^\infty([0, T], \mathbb{R}^2)$ with $\delta q(0) = (0, 0)^\top$ and $\delta q(T) = (0, 0)^\top$, that

$$X(\delta q) := I(q + \delta q) - I(q) = \tfrac{1}{2} \int_0^T \big(\|\delta \dot{q}(t)\|^2 - \delta q_2^2(t)\big) \, \mathrm{d}t \, .$$

(c) Next evaluate $X(\delta q)$ for a variation $\delta q = (\delta q_1, \delta q_2)^\top$ with

$$\delta q_2(t) = \sum_{n=1}^{\infty} c_n \, \sin\left(\tfrac{\pi n t}{T}\right) \quad , \quad \delta q_1(t) = 0 \quad , \text{where} \quad \sum_{n=1}^{\infty} (n\, c_n)^2 < \infty \, .$$

(d) Based on the result from (c), explain that $I(q)$ is minimal for times $0 < T < \pi$, but that for $l\pi < T < (l+1)\pi$, there is an l-dimensional subspace on which X is negative definite. ◊

The Challenge by Johann Bernoulli

In the *Acta Eruditorum*, which was the first scientific journal in Germany, Johann Bernoulli from Basel (at the time living in Groningen) published the following question in 1696:

> "If in a vertical plane two points A and B are given, assign to a moving point M such a path AMB along which the point M will pass most rapidly from A to B, based on its own gravity."

This problem was solved by many of the most significant contemporary mathematicians: next to Johann Bernoulli himself, by his older brother Jakob (whom he despised), by Leibniz, de l'Hospital, Tschirnhaus, and Newton.

The problem itself had already been mentioned by Galilei in [Gal2], where he claimed circular segments to be the solutions. With the newly developed methods of the differential calculus, segments of cycloids could now be identified as the true solution paths (Figure 8.3.1).
Despite certain variational problems like the one by Dido being known since antiquity, this *brachystochrone* problem (as it came to be known) is the beginning of the modern calculus of variations.
The quest is for a function $Y : [0, x_B] \to \mathbb{R}$, whose graph is the path of minimal time $T(Y)$. The path starts at the origin $A = 0 \in \mathbb{R}^2$ of the plane, and the end point has coordinates $B = (x_B, y_B)$ with $x_B > 0 > y_B$. Therefore $Y(0) = 0$ and $Y(x_B) = y_B$.
As M starts at A with zero speed, M will, under the influence of gravitation, have at height $y \le 0$ the speed $\sqrt{-2gy}$, where the acceleration of gravity is g. This makes

$$T(Y) = \int_0^{x_B} \sqrt{\tfrac{1+Y'(x)^2}{-2gY(x)}}\, \mathrm{d}x \,. \tag{8.3.6}$$

Instead of determining the Euler-Lagrange equation for $L(y, v) := \sqrt{\frac{1+v^2}{-2gy}}$, we write down the Hamilton function H that is associated with the Lagrange function L by Theorem 8.6, as a function of y and v:

$$H(y,v) = \mathrm{D}_2 L(y,v)\,v - L(y,v) = \frac{v^2}{\sqrt{(1+v^2)(-2gy)}} - \sqrt{\frac{1+v^2}{-2gy}} = \frac{-1}{\sqrt{(1+v^2)(-2gy)}}.$$

Denoting the constant value of $H < 0$ as $-1/\sqrt{4gr}$, we conclude therefore

$$1 + Y'(x)^2 = \frac{2r}{-Y(x)} \quad \text{or} \quad Y'(x) = \sqrt{-1 - \frac{2r}{Y(x)}}\,. \tag{8.3.7}$$

This is the differential equation of cycloids, i.e., of curves that have the parametric form

$$x(t) = r\,\big(t - \sin(t)\big) \quad , \quad y(t) = -r\,\big(1 - \cos(t)\big) \qquad (t \in \mathbb{R})$$

and thus satisfy $Y'(x) = \frac{y'(t)}{x'(t)} = \frac{-\sin(t)}{1-\cos(t)}$. (See the figure on page 169.)
The given solutions Y have a derivative that diverges at $x = 0$, so strictly speaking they do not belong to the class of admissible functions. On the other hand, nowadays even less smooth solutions are considered. Minimality and uniqueness of solutions can be shown by means of a maximum principle, which is a technique from the theory of optimal control. See SUSSMANN and WILLEMS [SW].
Johann Bernoulli claimed that his solution would be "quite useful also for other areas of science than mechanics". Well: The halfpipes used in skateboarding are often shaped with a cycloid cross section.
While the parabola shaped slides depicted in the beginning of the chapter are not optimal for time, they are more comfortable than the brachystochrone, since with them, the trip does not start with a free fall.

8.21 Exercise (**Tautochrone Problem**)
Show that the cycloid also solves the *tautochrone* problem, namely that all mass points starting at an arbitrary location of the curve with speed zero should arrive at the lowest point of the curve at the same time. ◊

8.4 Geodesic Motion

We investigate geodesic motion on submanifolds $M \subseteq \mathbb{R}^n$. Submanifolds are often defined as pre-images, and by their Definition 2.34, they can always be so represented locally. So let $W \subseteq \mathbb{R}^n$ be open, $F \in C^\infty(W, \mathbb{R}^m)$, and $0 \in F(W) \subset \mathbb{R}^m$ a regular value of F, and let $M := F^{-1}(0)$ be the submanifold. We are interested in the free motion on W, given by the Lagrangian

$$\tilde{L} : W \times \mathbb{R}^n_v \to \mathbb{R} \quad , \quad \tilde{L}(q,v) := \tfrac{1}{2}\|v\|^2$$

and its restriction by the holonomic constraint $q \in M$.

As on page 27, we restrict the coordinate diffeomorphism $\varphi : U \to \mathbb{R}^n$ to the neighborhood $V := U \cap M$ of q in M and write it in the form

$$\varphi\restriction_V(z) = \big(\psi(z), 0\big) \in \mathbb{R}^d \times \mathbb{R}^m ,$$

where $d := n - m$. The mapping $\psi : V \to \mathbb{R}^d$ can be smoothly inverted on its image $V' := \psi(V) \subset \mathbb{R}^d \times \{0\} \cong \mathbb{R}^d$, and we get the local coordinates

$$q := \psi^{-1} : V' \to V$$

in $V \subset M$, see Figure 8.4.1.

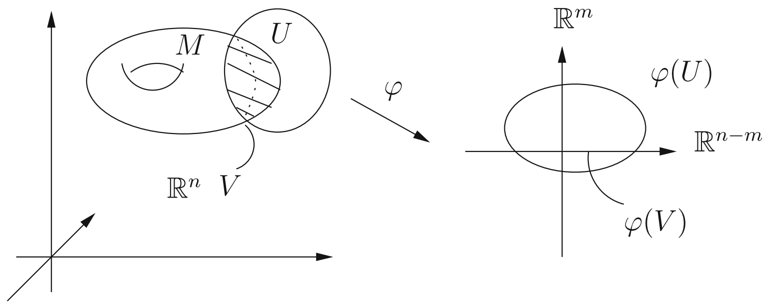


Figure 8.4.1 Construction of local coordinates in $V \subset M$

The Lagrangian L defined by

$$L : V' \times \mathbb{R}^d \to \mathbb{R} \quad , \quad L(x, w) := \tilde{L}\big(q(x), \mathrm{D}q(x)\, w\big) ,$$

namely $L(x, w) = \frac{1}{2}\, \|\mathrm{D}q(x)\, w\|^2$, is of the form

$$L(x, w) = \tfrac{1}{2} \sum_{i,j=1}^{d} g_{i,j}(x)\, w_i\, w_j \quad \text{with} \quad g_{i,j}(x) := \sum_{k=1}^{n} \frac{\partial q_k}{\partial x_i}(x)\, \frac{\partial q_k}{\partial x_j}(x) . \qquad (8.4.1)$$

8.22 Lemma *The matrix valued function $g : V' \to \mathrm{M}at(d, \mathbb{R})$ (also known as the* **first fundamental form***) is a* **Riemannian metric** *on V, i.e.,*

$$g_{i,j} = g_{j,i} \quad (i, j = 1, \ldots, d) \quad \textit{and} \quad g(x) > 0 \qquad (x \in V').$$

Proof:
- The symmetry of g follows immediately from its definition.
- For all $x \in V'$,

$$\sum_{i,j=1}^{d} g_{i,j}(x)\, w_i\, w_j = \|\mathrm{D}q(x)\, w\|^2 > 0 \qquad \big(w \in \mathbb{R}^d \setminus \{0\}\big);$$

here g is positive definite because, along with ψ, its inverse $q : V' \to V$ is a diffeomorphism as well, and therefore $\operatorname{rank}\big(\mathrm{D}q(x)\big) = d$. □

Since M can be covered by open sets of the form V, we obtain a metric on all of M. The Riemannian metric allows us to measure lengths of curves independently of a coordinate system. The length of a curve $c : [0, 1] \to V$ is defined by

$$\mathcal{L}(c) := \int_0^1 \sqrt{\textstyle\sum_{i,j}\, g_{i,j}\big(\tilde{c}(t)\big)\, \frac{\mathrm{d}\tilde{c}_i}{\mathrm{d}t}(t)\, \frac{\mathrm{d}\tilde{c}_j}{\mathrm{d}t}(t)}\, \mathrm{d}t = \int_0^1 \sqrt{2L\left(\tilde{c}(t),\, \frac{\mathrm{d}\tilde{c}}{\mathrm{d}t}(t)\right)}\, \mathrm{d}t$$

with $\tilde{c} := \psi \circ c$.[6]

The extremals of this functional, which is by (8.3.5) independent of the parametrization, coincide, up to parametrization, with the extremals of the functional given by the Lagrangian L.

By

$$\mathrm{D}_1 L(x, w) = \left(\tfrac{1}{2}\sum_{i,j=1}^{d} \frac{\partial g_{i,j}(x)}{\partial x_1} w_i w_j, \ldots, \tfrac{1}{2}\sum_{i,j=1}^{d} \frac{\partial g_{i,j}(x)}{\partial x_d} w_i w_j\right)$$

and

$$\mathrm{D}_2 L(x, w) = \left(\sum_{j=1}^{d} g_{1,j}(x) w_j, \ldots, \sum_{j=1}^{d} g_{d,j}(x) w_j\right),$$

the Lagrange equation is

$$\tfrac{1}{2}\sum_{i,j=1}^{d} \frac{\partial g_{i,j}(x)}{\partial x_k} \dot{x}_i \dot{x}_j = \sum_{j=1}^{d}\left(g_{k,j}(x)\ddot{x}_j + \sum_{i=1}^{d} \frac{\partial g_{k,j}}{\partial x_i} \dot{x}_j \dot{x}_i\right) \qquad (k = 1, \ldots, d).$$

We define the *Christoffel symbols* by

$$\Gamma^h_{i,j}(x) := \tfrac{1}{2} g^{h,k}(x) \left(\frac{\partial g_{k,j}}{\partial x_i}(x) + \frac{\partial g_{i,k}}{\partial x_j}(x) - \frac{\partial g_{i,j}}{\partial x_k}(x)\right) \qquad (i, j, h = 1, \ldots, d), \tag{8.4.2}$$

where $(g^{h,k})_{h,k=1}^d$ is the matrix inverse to $(g_{k,i})_{k,i=1}^d$, and the **Einstein summation convention** has been used, namely the convention that summation over indices occurring twice is to be tacitly understood.

[6] The length functional $\mathcal{L}$ is not to be confused with the Lagrangian L !

With this, we obtain the equations

$$\boxed{\ddot{x}_h + \Gamma^h_{i,j}(x)\dot{x}_i\dot{x}_j = 0 \qquad (h = 1, \ldots, d)\,.} \tag{8.4.3}$$

They are called the *geodesic equations*, and its solutions $t \mapsto x(t)$, *geodesics*.

We first notice that the metric tensor g, and thus also the geodesic equation, contain only informations about the *intrinsic* geometry of the d-dimensional surface M, independent of its (isometric) *imbedding*. So we can study geodesic motion on Riemannian manifolds (i.e., manifolds with a metric tensor) independent of any imbedding.

8.23 Exercise (**Euler-Lagrange Equation in Polar Coordinates**)
Consider the plane $\mathbb{R}^2$ with polar coordinates.

(a) Calculate the Christoffel symbols determined by the Lagrangian of free motion.
(b) Write down the length functional for curves in $\mathbb{R}^2$ and the Euler-Lagrange equations for straight lines in the plane, in polar coordinates. ◊

While the form of the matrix elements $g_{i,h}$ depends on the choice of a coordinate system, the following quantities can nevertheless be defined without coordinates.

- In Riemannian geometry, there exist invariants that characterize the *(intrinsic)* curvature of the space M.
- In addition, there are curvature invariants that characterize the form of the *imbedding* of a submanifold $M \subset \mathbb{R}^n$.

It is important to distinguish these two aspects.

8.24 Example (**Gauss Curvature**) A two dimensional surface $M \subset \mathbb{R}^3$ has, at each point $q \in M$, the intrinsic curvature $K(q) := k_1(q)\, k_2(q)$, which is the product of the reciprocals k_i of the curvature radii of a certain two curves. These curves are obtained as the intersection of M each with a two dimensional affine space through q that contains the normal to the surface. These two affine spaces are to be chosen in such a way that the radii of curvature become extremal.

The sign of the *principal curvature* k_i is positive, when the curve is curving towards the surface normal, otherwise negative. So whereas k_1 and k_2 depend on the choice of the normal, their product K, the *Gauss curvature*, is independent of it[7] (and therefore also well-defined for a non-orientable surface).

- A sheet of paper in $\mathbb{R}^3$ has intrinsic curvature $K = 0$, even though we can bend it. Even in the bent sheet of paper, one of the two principal curvatures k_i will vanish everywhere (see Figure 8.4.2).
- The sphere $\{x \in \mathbb{R}^3 \mid \|x\| = R\}$ of radius R: Its curvature is $K = k_1\, k_2 = R^{-2} > 0$. The Gauss curvature, say, of a soft contact lens will also not change if it gets compressed in one direction.
- Saddle: Here the signs of k_1 and k_2 differ, hence $K < 0$. ◊

[7] According to the *Theorema Egregium* by Gauss it can be written in terms of the metric g, its first and second derivatives. So it is an intrinsic quantity.

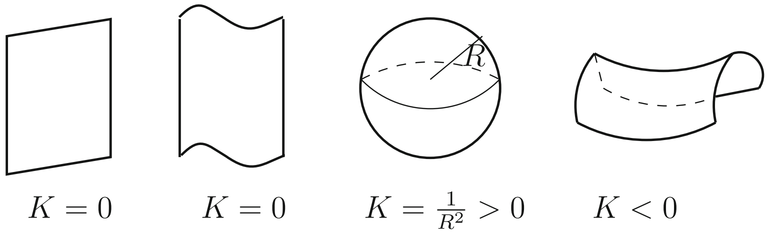


Figure 8.4.2 Gauss curvature K of surfaces

From the derivation of the Lagrangian L, it transpires that we can interpret

$$\sqrt{2L(x,\dot{x})} = \sqrt{\sum_{i,j=1}^{d} g_{i,j}(x)\,\dot{x}_i\dot{x}_j}$$

as the speed (absolute value of the velocity) of a particle.

As can be seen by plugging the geodesic equations into the time derivative $\frac{\mathrm{d}}{\mathrm{d}t}L = \frac{\partial L}{\partial x}\dot{x} + \frac{\partial L}{\partial \dot{x}}\ddot{x}$ of the Lagrangian, this time derivative is zero, hence L is a constant of motion. So the speed of the particle on M will be constant.

A particularly simple class of imbedded Riemannian manifolds are the surfaces of revolution.

8.25 Definition *Let I be an interval and $R \in C^\infty(I, \mathbb{R}^+)$. Then,*

$$M := \left\{x \in \mathbb{R}^3 \mid x_3 \in I, x_1^2 + x_2^2 = (R(x_3))^2\right\}$$

is called the **surface of revolution** *with* **profile** *R.*

The simplest surface of revolution is the cylinder (with constant R), see Figure 8.4.3.

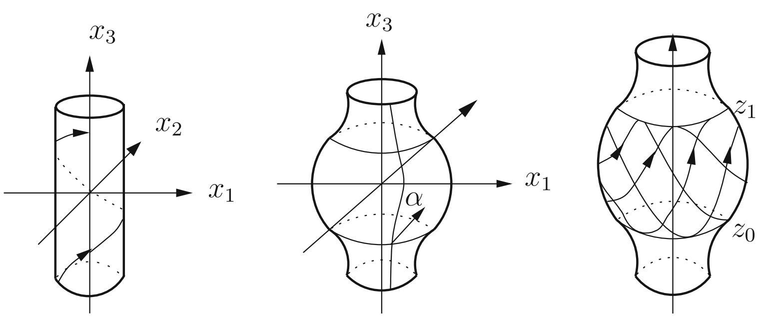


Figure 8.4.3 Left: Geodesic on a cylinder. Center: Definition of the angle α. Right: Geodesic on a surface of revolution

We parametrize a surface of revolution by $z \in I$ and φ with

$$x_1 = R(z)\cos(\varphi) \quad , \quad x_2 = R(z)\sin(\varphi) \quad , \quad x_3 = z\,.$$

Thus (8.4.1) yields the metric tensor $g(z,\varphi) = \begin{pmatrix} 1+(R'(z))^2 & 0 \\ 0 & R^2(z) \end{pmatrix}$ and, with the velocity components w_z and w_φ in z and φ directions respectively, the Lagrangian

$$L(z,\varphi,w_z,w_\varphi) = \tfrac{1}{2}\left[\left(1+(R'(z))^2\right) w_z^2 + R^2(z) w_\varphi^2\right].$$

8.26 Exercise (**Geodesics on Surface of Revolution**)

(a) Show by calculating the Christoffel symbols that the geodesic equations on M are

$$\ddot{\varphi} + \tfrac{2R'(z)}{R(z)}\dot{\varphi}\dot{z} = 0 \quad , \quad \ddot{z} - \tfrac{R(z)R'(z)}{R'(z)^2+1}\dot{\varphi}^2 + \tfrac{R'(z)R''(z)}{R'(z)^2+1}\dot{z}^2 = 0\,.$$

(b) Show that the *meridians* (in other words curves in M with constant angle $\varphi \equiv c \in \mathbb{R}$), parametrized by arclength, are geodesics in M.

(c) Which *parallels of latitude*, i.e., curves with constant $z \equiv c \in \mathbb{R}$, are geodesics? ◊

The momentum p_φ conjugate to φ is given by $p_\varphi = \frac{\partial L}{\partial w_\varphi} = R^2(z)\, w_\varphi$, and $p_z = \frac{\partial L}{\partial w_z} = \left(1+(R'(z))^2\right) w_z$. Therefore the Hamiltonian H related to L by the Legendre transform is

$$\begin{aligned} H(p_z,p_\varphi,z,\varphi) &= p_z w_z + p_\varphi w_\varphi - L(z,\varphi,w_z,w_\varphi) \\ &= L\big(z,\varphi,w_z(p_z,z),w_\varphi(p_\varphi,z)\big) = \tfrac{1}{2}\left(\frac{p_z^2}{1+(R'(z))^2} + \frac{p_\varphi^2}{R^2(z)}\right). \end{aligned}$$

The equations of motion are

$$\dot{p}_z = -\frac{\partial H}{\partial z} = \frac{R'(z)R''(z)p_z^2}{(1+(R'(z))^2)^2} - R^{-3}(z)p_\varphi^2 \quad , \quad \dot{p}_\varphi = -\frac{\partial H}{\partial \varphi} = 0 \quad ,$$

$$\dot{z} = \frac{\partial H}{\partial p_z} = \frac{p_z}{1+(R'(z))^2} \quad \text{and} \quad \dot{\varphi} = \frac{\partial H}{\partial p_\varphi} = \frac{p_\varphi}{R^2(z)}\,.$$

It is immediate that p_φ is a constant of motion.[8]

If we denote by α the angle of the direction of a geodesic with a meridian at the same point (see Figure 8.4.3), we obtain

8.27 Theorem (**Clairaut**)
For a geodesic on a surface of revolution, $R(z)\ \sin\alpha$ is constant in time.

[8]The system is integrable in the sense of the definition on page 330.
We can also understand the fact that p_φ is constant by noting that p_φ is the 3-component of the angular momentum. It is conserved because the surface of revolution is invariant under rotation about the 3-axis: see Noether's theorem on page 351.

Proof: We have $R(z)\,\dot{\varphi} = \|v\|\sin\alpha$ with speed $\|v\| = \sqrt{2L}$. Therefore $p_\varphi = R^2\,\dot{\varphi} = R\,\|v\|\,\sin\alpha$, and as $\|v\|$ and p_φ are constant, we get that $R\sin\alpha = \text{const}$. □

As $|\sin\alpha| \le 1$, a given (nonzero) value of the Clairaut constant implies that R cannot become too small.

This could imply that motion along a certain geodesic is confined to a band $z_0 \le z \le z_1$ with $R(z_0) = R(z_1) = |\text{const}|$, see the figure on the right of 8.4.3.

8.28 Literature
An introduction to differential geometry can be found in Klingenberg [Kli1].

A deeper introduction to Riemannian geometry is offered by the books [GHL] by Gallot, Hulin and Lafontaine, [Kli2] by Klingenberg, and [Pat] by G. Paternain, where the latter two have a focus on geodesic flow.

[Berg] by Berger gives a survey without proofs. [JLJ] by Jost and Li- Jost is a monograph on the calculus of variations. In [Ra], Paul Rabinowitz gives an overview of variational methods for Hamiltonian systems. ◊

8.5 The Jacobi Metric

An immediate application of (semi-)Riemannian geometry to mechanics is the geodesic motion within metrics of general relativity, for example the metric of a black hole. But its relevance to mechanics is not limited to this application.

Many problems of mechanics are described by motion within a potential. So it would be advantageous to be able to generalize the theory of geodesic motion, which is rich and comparatively well worked out, to such problems with potential whose Hamilton function is $H(p,q) = \frac{\|p\|^2}{2m} + V(q)$. To some extent, this is possible, as was understood by Jacobi and others in the 19th century.

The reason for this relation is in the following theorem.

8.29 Theorem *Let $H_1, H_2 \in C^2(M,\mathbb{R})$ on the phase space $M := \mathbb{R}^n_p \times U$ ($U \subseteq \mathbb{R}^n_q$ open), and suppose for the regular values $h_i \in H_i(M)$, $i = 1, 2$ that*

$$\Sigma := H_1^{-1}(h_1) = H_2^{-1}(h_2)\,.$$

Then the maximal Hamiltonian flows generated by H_1 and H_2 respectively on the hypersurface Σ have the same family of orbits.

Proof: $\Sigma \subset M$ is a $(2n-1)$-dimensional submanifold, since the $h_i \in H_i(M)$ are regular. In (p,q) coordinates, the Hamiltonian vector fields generated by H_i are of the form (Figure 8.5.1)

$$X_{H_i} = J\ \nabla H_i \quad \text{, with} \quad \mathbb{J} = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}.$$

At each point on the energy shell Σ, both ∇H_1 and ∇H_2 are orthogonal to the $(2n-1)$-dimensional tangent plane, because the energy shell is an equipotential surface of both functions. Therefore ∇H_1 and ∇H_2 are parallel. Since h_1 and h_2 are regular values of H_1 and H_2 respectively, $\nabla H_i(x) \neq 0$ for all $x \in \Sigma$. Hence

$$\nabla H_1(x) = f(x)\ \nabla H_2(x) \qquad (x \in \Sigma),$$

and $f : \Sigma \to \mathbb{R}$ is smooth and not zero. This implies that on the level set,

$$X_{H_1}(x) = f(x)\ X_{H_2}(x) \qquad (x \in \Sigma),$$

so the Hamiltonian vector fields are parallel (or anti-parallel) and do not vanish. The statement follows from this. □

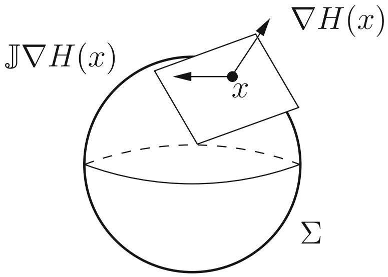


Figure 8.5.1 Gradient and Hamiltonian vector field

Therefore the shape of the energy shell alone determines the orbits of the Hamiltonian flow. On the other hand, the parametrization by time is not so determined; in the situation of the theorem, the reparametrization is obtained by integrating f along orbits.

We now apply Theorem 8.29 to motion in a potential.

8.30 Definition *Let $V \in C^2(\mathbb{R}^n_q, \mathbb{R})$, $h \in \mathbb{R}$, and $U \subseteq \mathbb{R}^n_q$ open. If $V\restriction_U < h$, we call the metric g_h on U with components*

$$\big(g_h(q)\big)_{i,j} := \big(h - V(q)\big) \cdot \delta_{i,j} \qquad (i, j = 1, \ldots, n)$$

the **Jacobi metric** *on U for energy h.*

The Lagrangian for geodesic motion in this metric is

$$L(q, v) = \tfrac{1}{2}\textstyle\sum_{i,j=1}^n \big(g_h(q)\big)_{i,j}\ v_i\, v_j\ .$$

The canonically conjugate momentum is $p = \frac{\partial L}{\partial v}$, hence $p(q, v) = g_h(q)\, v$. Thus the Hamiltonian of the geodesic motion equals

$$H_2 : \mathbb{R}^n_p \times U \to \mathbb{R} \quad , \quad H_2(p, q) := \frac{1}{2(h - V(q))}\ \|p\|^2\ .$$

We compare it with the Hamiltonian

$$H : \mathbb{R}^n_p \times \mathbb{R}^n_q \to \mathbb{R} \quad , \quad H(p, q) := \tfrac{1}{2}\|p\|^2 + V(q)\,. \tag{8.5.1}$$

8.31 Theorem
For a solution $(p, q) : I \to \mathbb{R}^n_p \times U$ to the Hamiltonian equations of (8.5.1) with energy $h := H(p, q)$, there exists a diffeomorphism $\varphi : J \to I$ such that $q \circ \varphi : J \to U$ solves the geodesic equation (8.4.3) for the Jacobi metric on U.

Proof:
• Restricted to the phase space over U, one has $H^{-1}(h) = H_2^{-1}(1)$.
• Transforming the Hamiltonian equations of H_2 into a system of second order differential equations in the coordinates $(q_1, \ldots, q_n)$, one obtains the geodesic equations for the Jacobi metric.
• The above Theorem 8.29 (with $H_1 := H$) now yields the claim. □

8.32 Example (Double Pendulum)
A typical application of the Jacobi metric is given by the double pendulum. We study the case when a rod of length R_1, attached to the origin, oscillates in the q_1-q_3 plane, and another rod of length R_2, attached to the end of the former rod, oscillates in the plane determined by the first rod and the q_2-axis. So the planes of oscillation are orthogonal to each other.[9]

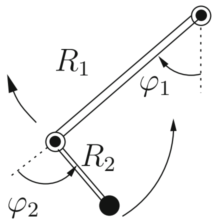
Assuming $R_1 > R_2 > 0$, the set of positions $q : S^1 \times S^1 \to \mathbb{R}^3$,

$$\begin{aligned} q(\varphi_1, \varphi_2) &= \begin{pmatrix} \cos\varphi_1 & 0 & -\sin\varphi_1 \\ 0 & 1 & 0 \\ \sin\varphi_1 & 0 & \cos\varphi_1 \end{pmatrix} \left(R_1 \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} + R_2 \begin{pmatrix} 0 \\ \sin\varphi_2 \\ -\cos\varphi_2 \end{pmatrix} \right) \\ &= R_1 \begin{pmatrix} \sin\varphi_1 \\ 0 \\ -\cos\varphi_1 \end{pmatrix} + R_2 \begin{pmatrix} \sin\varphi_1 \cos\varphi_2 \\ \sin\varphi_2 \\ -\cos\varphi_1\cos\varphi_2 \end{pmatrix} \end{aligned}$$

of the mass point is a two dimensional submanifold of $\mathbb{R}^3$. More specifically, it is diffeomorphic to the 2-torus $\mathbb{T}^2$, where for $n \in \mathbb{N}$, the *n-torus* is defined by

$$\mathbb{T}^n := \times_{k=1}^n S^1 \quad \text{with} \quad S^1 = \{x \in \mathbb{C} \mid |x| = 1\}$$

(see figure to the right). The Lagrangian



Double pendulum

$$\tilde{L}(q, v) := \tfrac{1}{2}\|v\|^2 - \tilde{V}(q) \quad \text{with} \quad \tilde{V}(q) := g q_3$$

($g > 0$ acceleration of gravity) attains, after plugging in the holonomic constraints, the form

[9]The discussion in [Ar2], §45 for the *planar* double pendulum is problematic because the metric on the configuration space becomes degenerate in this case.

$$L(\varphi, w) = \tfrac{1}{2}\left((R_1 + R_2\cos(\varphi_2))^2 w_1^2 + R_2^2 w_2^2\right) - V(\varphi)\,,$$

with potential

$$V(\varphi) = -g\cos(\varphi_1)(R_1 + R_2\cos(\varphi_2)). \tag{8.5.2}$$

For arbitrary smooth potentials V, we thus obtain $H : \mathbb{R}_p^2 \times \mathbb{T}^2 \to \mathbb{R}$ of the form

$$H(p, \varphi) = \tfrac{1}{2}\left((R_1 + R_2\cos(\varphi_2))^{-2} p_1^2 + R_2^{-2} p_2^2\right) + V(\varphi)\,.$$

As can be seen from the numerical solution of the Hamiltonian differential equation

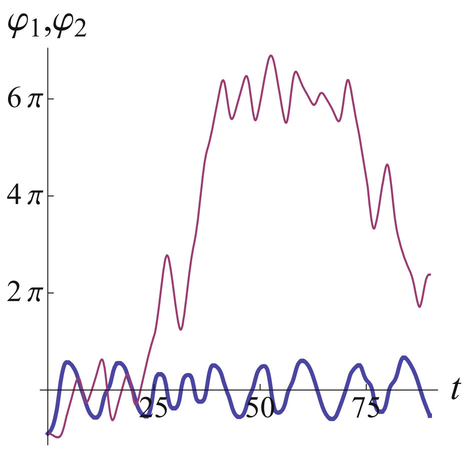


Figure 8.5.2 Numerical solution of the differential equation of the double pendulum (the angle drawn in bold is φ_1)

for the potential (8.5.2) in Figure 8.5.2, the dynamics of the double pendulum is complicated. Nevertheless, by means of the Jacobi metric, one can prove the existence of certain periodic orbits.

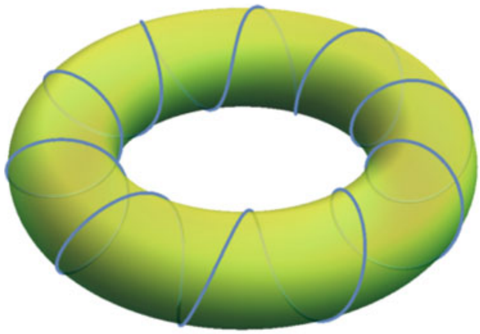
For the total energy $h > \max_{\mathbb{T}^2} V$ (which means in the case of the gravitation potential (8.5.2) that $h > g \cdot (R_1 + R_2)$), the Jacobi metric can be represented by the following positive definite symmetric matrix, which is defined on all of $\mathbb{T}^2$:

$$g_h(\varphi) = \big(h - V(\varphi)\big)\begin{pmatrix}(R_1 + R_2\cos(\varphi_2))^2 & 0\\ 0 & R_2^2\end{pmatrix} \qquad (\varphi \in \mathbb{T}^2).$$

Similar to the discussion above, we can study the geodesic motion on $\mathbb{T}^2$ with respect to the Jacobi metric. $\diamondsuit$

8.33 Theorem *For energy $h > \max_{\varphi\in\mathbb{T}^2} V(\varphi)$ and $(m, n) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$, there exists a periodic motion with this total energy for which the first segment of the double pendulum rotates m times and the second n times.*

Proof: On $\mathbb{T}^2$, we look for a closed geodesic with respect to the Jacobi metric, say in the form sketched in the figure on the right.



Closed geodesic on a torus

Such a geodesic exists by general principles from the calculus of variations (or by Morse theory) for each nontrivial class of paths in the homotopy group $\pi_1(\mathbb{T}^2) \cong \mathbb{Z}^2$ (see Theorem G.24).

The idea in this argument is to reduce the length of a given curve with the prescribed numbers of rotations as much as possible. The result will be a closed geodesic with those numbers of rotations, because these numbers do not change under shortening. In analogy to Theorem 8.31, there is a periodic solution for the double pendulum corresponding to the closed geodesic; a solution that differs from the geodesic only by the parametrization with respect to time. □

The *Maupertuis principle*, which claims that the action integral $\int_\gamma p \cdot dq$ on the energy surface is extremal for the trajectories (see for instance §45D in [Ar2]), is closely related to the Jacobi metric. Maupertuis viewed his principle as a proof of God, his successors viewed it as a mathematical theorem.

8.6 Fermat's Principle

The *refractive index n* of a transparent medium is the ratio between the phase velocity of the light in vacuum (i.e., the speed of light) and the phase velocity in the medium. For instance one finds in the relevant tables that for air $n = 1.000\,292$, and for water $n = 1.33$.

As a matter of fact, n in general varies with the density (for example the refractive index of the atmosphere tends to 1 for large elevations) and it also varies with the frequency of the light; this latter fact is being used in prisms.

Let us first look at monochromatic (single-colored) light. Then n is a real function depending on location. We assume $n \in C^2(U, \mathbb{R})$, for an open set $U \subseteq \mathbb{R}^d$. *Fermat's principle* says that the trajectory of a ray of light $c : I \to U$ in a medium, parametrized over an interval I of time, is described by the Lagrangian of the *optical distance*

$$\tilde{L} : U \times \mathbb{R}^d_v \to \mathbb{R} \quad , \quad (q, v) \mapsto n(q)\|v\| ,$$

namely the length element of a metric that differs from the Euclidean metric by a location dependent factor.

As seen already in Section 8.4, we transition to the Lagrangian

$$L : U \times \mathbb{R}^d_v \to \mathbb{R} \quad , \quad (q, v) \mapsto \tfrac{1}{2}\bigl(n(q)\|v\|\bigr)^2$$

of geodesic motion. Then the Lagrange equations read

$$n(q)\ddot{q} = \|\dot{q}\|^2\, \nabla n(q) - 2\langle \nabla n(q), \dot{q}\rangle\, \dot{q}\,, \tag{8.6.1}$$

and one can check by direct calculation that L is constant along solution curves.

8.34 Example (**Mirage**)
We assume that the density of air is a function of the elevation above ground. Then the horizontal direction of the ray is constant in time, and we calculate in the plane $U := \mathbb{R}^2 \subset \mathbb{R}^3$ parametrized by (x, y) and spanned by this horizontal direction and the vertical direction. Here y measures the elevation, and we write the refractive index as a function $y \mapsto n(y)$ of this elevation.

It is useful, if possible, to parametrize the ray by the horizontal component $x \in \mathbb{R}$ of the location, rather than by time. Then the vertical component $x \mapsto y(x)$ satisfies the differential equation

$$n(y)y'' = n'(y)\bigl(1 + (y')^2\bigr) \tag{8.6.2}$$

following from (8.6.1). Let us now assume that n is an affine function, i.e., it is of the form

$$n(y) = n_0 + ky\,; \tag{8.6.3}$$

then, for $k \neq 0$, one obtains the following solutions, dependent on the parameters c and x_0:

$$y(x) = \frac{\cosh\bigl(c(x - x_0)\bigr)}{c} - \frac{n_0}{k}\,. \tag{8.6.4}$$

Choosing now $x_0 := \operatorname{arcosh}(cn_0/k)/c$, we get a one-parameter family of solutions y_c satisfying $y_c(0) = 0$. So the rays all emanate from the origin, but will cross a second time (see Figure 8.6.1).

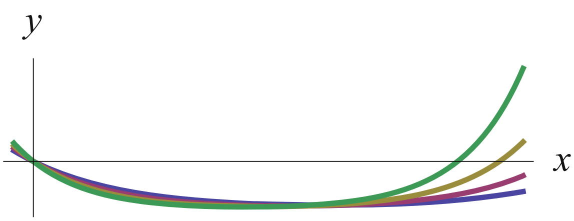


Figure 8.6.1 Mirage: rays emanating from one point

This means that for large distances x, the object will be reflected, i.e., appear to be upside down. If the air is hotter near the ground, hence $k > 0$, one does not see objects close to the ground at a far distance.

The ansatz (8.6.3) is not entirely true to reality. As a matter of fact, the refractive index will become approximately constant again for larger y, and we will see objects twice (once directly, once reflected), see Figure 8.6.2. ◊

Figure 8.6.2 Mirage over the Mojave desert (Photo: Wikipedia, https://es.wikipedia.org/wiki/Archivo:Desert_mirage_62907.JPG)

8.35 Exercises (Refraction (of Light))

1. Derive the Lagrange equations (8.6.1) and from these, the ODE (8.6.2).
2. How does one find the solution (8.6.4) by using that the Lagrange function is constant?
3. Let the x-axis represent the boundary between two media with different speeds of light. Above the x-axis, let the speed of light be $c_1 > 0$, and below $c_2 > 0$. Now consider the path of a ray of light that starts from the point $a_1 = (0, y_1)$ with $y_1 > 0$, undergoes refraction in the x-axis, and then meets the point $a_2 = (x_2, y_2)$ with $y_2 < 0$. By Fermat's principle, the point of refraction $a_0 = (x_0, 0)$ is located in such a way that the travel time

$$T(a_0) := \frac{\|a_1 - a_0\|}{c_1} + \frac{\|a_2 - a_0\|}{c_2}$$

of the ray is minimal. Show for this point a_0, *Snell's law of refraction* [10]

$$c_2 \sin \alpha_1 = c_1 \sin \alpha_2$$

(with the angles α_1, α_2 of the segments against the normal).
Hint: It is not necessary to calculate x_0 explicitly. ◊

[10]Named after the Dutch mathematician *Willebrord van Roijen Snell* (1580–1626). The law was first found by the Arabic mathematician *Abu Sa'd Ibn Sahl* (ca. 940–1000), in the form $n_1 \sin \alpha_1 = n_2 \sin \alpha_2$, with the refractive indices n_1, n_2 in the two media.

8.7 Geometrical Optics

Lenses have two refractive surfaces. If $r > 0$ is the radius of the lens and $D_r := \{(y,z) \in \mathbb{R}^2 \mid y^2 + z^2 \le r^2\}$ denotes the disc of radius r, we describe the refractive surfaces as graphs of smooth functions

$$O : D_r \to \mathbb{R}_x .$$

We assume that $O(-y,-z) = O(y,z)$ so that the surface is *symmetric* with respect to the *optical axis* $\{(x,y,z) \in \mathbb{R}^3 \mid y = z = 0\}$.

8.36 Examples (Refractive Surfaces)

1. If the surface is *planar*, then O is a constant function, whose value describes the position of the surface on the optical axis.
2. If the surface is *spherical*, with radius of curvature $R \in \mathbb{R}$, $|R| > r$, then

$$O(y,z) = O(0,0) + R\left(\sqrt{1 - \tfrac{y^2+z^2}{R^2}} - 1\right). \qquad \diamondsuit$$

Indexing the two surfaces in such a way that $O^+ > O^-$, we call either surface *convex* or *concave* respectively, if and only if the corresponding function O^- or $-O^+$ is convex or concave respectively. Lenses with convex surfaces gather (focus) the light (assuming the medium between the surfaces has the larger n).

A determining parameter for the properties of the lens is its focal length. To define it, we first note that due to the assumption of symmetry, a ray arriving along the optical axis will not be deflected by the lens.

Usually the focal length is defined as the distance along the optical axis between the lens and the focal point. This however fails to be precise, not only because lenses have a finite thickness, but also because spherical lenses do not diffract rays that are parallel to the optical axis into a single focus:

8.37 Example (Plano-convex Lens)
We consider a *plano-convex* spherical lens, i.e., we set

$$O^- := 0 \quad \text{and} \quad O^+(y,z) := d + R\left(\sqrt{1 - \frac{y^2+z^2}{R^2}} - 1\right),$$

with $d > 0$ and $R > 0$. Rays arriving from the left and parallel to the axis are not diffracted by the planar surface. Denoting by $\ell := \sqrt{y^2+z^2} < R/n$ the distance of the parallel ray to the optical axis, it will hit the right surface at an angle $\theta_1 := \arcsin(\ell/R)$ against the normal direction to the surface. By Snell's law of refraction, with refractive index $n > 1$ inside the lens, and 1 outside, one obtains that

$$\sin\theta_2 = n\sin\theta_1 = \frac{n\ell}{R}.$$

Thus the ray hits the optical axis at the point

$$X(\ell) = \tilde{O}(\ell) + \ell \cot(\theta_2 - \theta_1)$$

with $\tilde{O}(\ell) := d + R(\sqrt{1 - \ell^2/R^2} - 1)$. Taylor expansion in ℓ yields

$$X(\ell) = d + \frac{R}{n-1}\left(1 - \frac{1}{2}\left(\frac{n\ell}{R}\right)^2 + \mathcal{O}(\ell^4)\right).$$

The focus is defined by the leading term $X(0) = d + \frac{R}{n-1}$, which is independent of ℓ. ◊

The phenomenon that spherical lenses and mirrors focus rays parallel to the axis only approximately in one point is called *spherical aberration*. Rather than seeing a focus, one observes bright curves at the envelope of the bundle of rays, called caustics (see the left one of the Figure 8.7.1 on page 186).

In practical technology, one strives to avoid this phenomenon as much as possible by a combination of lenses, or by aspherical lenses. On the other hand, caustics are an interesting phenomenon occurring frequently in nature, see for instance Figure 8.7.1, right half.

8.38 Definition (**Caustic**)
Let $\iota : L \to P$ *be the imbedding of an* n*-dimensional submanifold into the phase space* $P := \mathbb{R}^n_p \times \mathbb{R}^n_q$*, and* $\pi : P \to \mathbb{R}^n_q$*,* $(p, q) \mapsto q$ *the projection of the configuration space* $\mathbb{R}^n_q$*.*
Then the set of singular values of $\pi \circ \iota : L \to \mathbb{R}^n_q$ *is called a* **caustic**.

8.39 Example (**Folding Singularity**) For $n = 1$ and the imbedding $\iota : \mathbb{R} \to P$, $\iota(x) = \big(x, (x-a)^2 + b\big)$, the caustic consists of the single point $b \in \mathbb{R}_q$. ◊

8.40 Literature In the example, one can see that small changes in the imbedding ι of L will only change the location of the caustic, but will not make it disappear. This is a general phenomenon, and it is possible to classify typical local shapes of caustics. The relevant theory is developed in the book [AGV] by ARNOL'D, GUSEIN-ZADE and VARCHENKO. In the Les-Houches lectures [Berr] by M. BERRY, many applications to physics can be found.

In applications to mechanics and optics, the submanifolds L that occur are frequently Lagrangian (see Chapter 10.4). ◊

8.41 Exercise (**Reflection of Light in a Cup**)
We consider the circle $S^1 \subset \mathbb{R}^2$ and a family of rays parallel to the 1-axis that, for the first time, hit the circle at $A(\varphi) := \binom{\sin\varphi}{\cos\varphi}$ with $\varphi \in [0, \pi]$.

There they will be reflected according to the law

"reflected angle = -- incoming angle".

(a) First show that the next reflection occurs at the point $A(3\varphi) \in S^1$.
(b) Let $B_t(\varphi) := tA(\varphi) + (1-t)A(3\varphi)$ $(t \in [0,1])$ be a point on the ray between the first and second reflection. For which value t_0 of t is it true (with $\mathbb{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$) that

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}\varphi} B_t(\varphi),\ \mathbb{J}\left(A(3\varphi) - A(\varphi)\right) \right\rangle = 0\ ?$$

(c) Calculate the *caustic* $B_{t_0} : [0, \pi] \to \mathbb{R}^2$, which is the blue curve in Figure 8.7.1 on the left. Which is the brightest point on the caustic? ◊

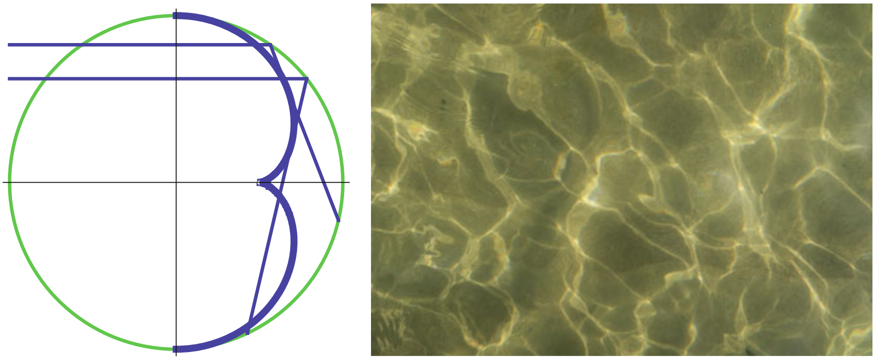


Figure 8.7.1 Caustic in a cup that is illuminated from the left (left figure). Caustics on the ocean floor (right figure)

Below however, we will study *linear optics*, in which the propagation of rays near the optical axis is described by the multiplication of matrices.

In doing so, we will describe rays as graphs of curves $q : \mathbb{R} \to \mathbb{R}^2$.

As long as they remain in a medium with constant refractive index n, one has, for an appropriate vector $p \in \mathbb{R}^2$,

$$q(x) = q(x_0) + \frac{x - x_0}{n} p\,,$$

where p parametrizes the direction of the ray and is independent of z. Thus

$$\begin{pmatrix} p(x) \\ q(x) \end{pmatrix} = M \begin{pmatrix} p(x_0) \\ q(x_0) \end{pmatrix} \quad \text{with} \quad M := \begin{pmatrix} 1\!\!1 & 0 \\ \frac{x-x_0}{n} 1\!\!1 & 1\!\!1 \end{pmatrix} \in \mathrm{Sp}(4, \mathbb{R})\,. \tag{8.7.1}$$

In the linear approximation, only the second degree Taylor polynomial needs to be considered near an interface. Due to the symmetry assumption $O(-q) = O(q)$, the graph of the interface is of the form

$$O(q) = O(0) + \tfrac{1}{2}\langle q, Aq\rangle + \mathcal{O}(\|q\|^4) \tag{8.7.2}$$

with a matrix $A \in \mathrm{M}at(2, \mathbb{R})$ that can be chosen to be symmetric.

8.42 Exercise (**Linear Optics**) Show that in linear approximation, the refraction in an interface described by (8.7.2) is given by the linear mapping

$$\begin{pmatrix} p \\ q \end{pmatrix} \mapsto N \begin{pmatrix} p \\ q \end{pmatrix} \quad \text{with} \quad N := \begin{pmatrix} 1\!\!1 & \Delta n\, A \\ 0 & 1\!\!1 \end{pmatrix} \in \mathrm{Sp}(4, \mathbb{R}) \;, \tag{8.7.3}$$

where Δn denotes the difference of the two refractive indices. ◊

So the calculation of the linearized propagation is achieved by multiplying matrices in $\mathrm{Sp}(4, \mathbb{R})$.

8.43 Example (**Spherical Lens**)
Let the refractive index of glass be $n > 1$, and approximate the refractive index of air by 1. Let the two radii of curvature be R^- and R^+, and the thickness of the lens $d > 0$. This lens is then described by the matrix $L(d) \equiv$

$$L := \begin{pmatrix} 1\!\!1 & \frac{(1-n)}{R^+} 1\!\!1 \\ 0 & 1\!\!1 \end{pmatrix} \begin{pmatrix} 1\!\!1 & 0 \\ \frac{d}{n} 1\!\!1 & 1\!\!1 \end{pmatrix} \begin{pmatrix} 1\!\!1 & \frac{n-1}{R^-} 1\!\!1 \\ 0 & 1\!\!1 \end{pmatrix} = \begin{pmatrix} \left(1+\frac{d(1-n)}{R^+ n}\right) 1\!\!1 & \frac{(1-n)(d(n-1)+n(R^- - R^+))}{nR^- R^+} 1\!\!1 \\ \frac{d}{n} 1\!\!1 & \left(1+\frac{d(n-1)}{nR^-}\right) 1\!\!1 \end{pmatrix}$$

If we set

$$\frac{1}{\Delta x} := (n-1)\left(\frac{1}{R^+} - \frac{1}{R^- + d\frac{n-1}{n}}\right),$$

then the 4×4-matrix ML for $M := \begin{pmatrix} 1\!\!1 & 0 \\ \Delta x 1\!\!1 & 1\!\!1 \end{pmatrix} \in \mathrm{Sp}(4, \mathbb{R})$ is of the form $ML = \begin{pmatrix} c_{11} 1\!\!1 & c_{12} 1\!\!1 \\ c_{21} 1\!\!1 & c_{22} 1\!\!1 \end{pmatrix}$ with $c_{22} = 0$ and $c_{21} = -1/c_{12}$. So Δx is the distance of the focus from the right surface of the lens, and for a thin lens, the focal distance f is approximately

$$\frac{1}{f} \approx (n-1)\left(\frac{1}{R^+} - \frac{1}{R^-}\right). \tag{8.7.4}$$

By definition, a *thin lens* is described by the matrix

$$N := \lim_{d \to 0} L(d) = \begin{pmatrix} 1\!\!1 & \frac{\Delta n}{R} 1\!\!1 \\ 0 & 1\!\!1 \end{pmatrix}$$

with an effective radius of curvature R and $\frac{1}{R} = \frac{1}{R^+} - \frac{1}{R^-}$.

We can also verify, by multiplying matrices, the well-known *lens equation* or *imaging equation*

$$\boxed{\frac{1}{f} = \frac{1}{b} + \frac{1}{g}\,,}$$

where b denotes the image distance and g the object distance, see Figure 8.7.2. Generally speaking, in optics, reciprocals of distances are more natural than the distances themselves, as (effectively) infinite distances often occur. ◊

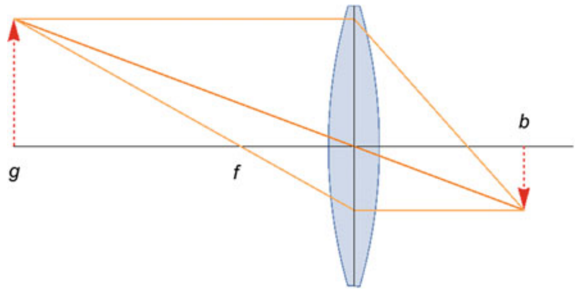


Figure 8.7.2 Imaging equation for a thin lens

It is now easy to develop optical instruments based on this linear approximation.

8.44 Example (**Kepler Telescope**)
We put two thin lenses with focal lengths $f_1, f_2 > 0$ in a row in such a way that their distance equals d. The entire assembly is therefore represented by the matrix

$$L_d := \begin{pmatrix} \mathbb{1} & -\mathbb{1}/f_2 \\ 0 & \mathbb{1} \end{pmatrix} \begin{pmatrix} \mathbb{1} & 0 \\ d\mathbb{1} & \mathbb{1} \end{pmatrix} \begin{pmatrix} \mathbb{1} & -\mathbb{1}/f_1 \\ 0 & \mathbb{1} \end{pmatrix} = \begin{pmatrix} (1-\frac{d}{f_2})\mathbb{1} & \frac{d-f_1-f_2}{f_1 f_2}\mathbb{1} \\ d\mathbb{1} & (1-\frac{d}{f_1})\mathbb{1} \end{pmatrix}. \qquad (8.7.5)$$

For relaxed viewing, the parallel rays from a star should again be parallel as they exit the eyepiece. For $d = f_1 + f_2$, this is satisfied because in this case, the top right entry of L_d vanishes: $L_d = \begin{pmatrix} -\frac{f_1}{f_2}\mathbb{1} & 0 \\ (f_1+f_2)\mathbb{1} & -\frac{f_2}{f_1}\mathbb{1} \end{pmatrix}$. Angles will then be magnified by a factor $-\frac{f_1}{f_2}$. As can be seen from the negative sign, the image is upside down.
Typical numerical values for an amateur telescope are about $f_1 = 1\,\mathrm{m}$ for the objective and $f_2 = 20\,\mathrm{mm}$ for the eyepiece, hence a 50-fold magnification. ◊

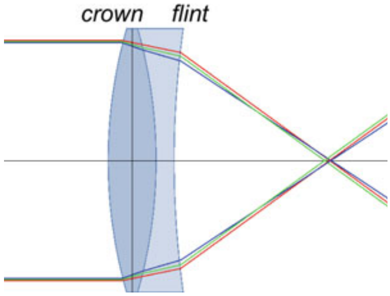
8.45 Exercises (**Optical Devices**)

1. Design a microscope consisting of two thin lenses.
2. Calculate the matrix N of a lens that, in linear approximation, corrects for a nearsighted eye with astigmatism. ◊

As a final practical application, we want to correct for the *chromatic aberration* of lenses. This aberration is caused by the fact that the refractive index of glass depends on the frequency of the light. $n(\omega)$ increases with the frequency ω, which results in the focal length of a collecting lens being smaller for blue light than for red light.

To correct this, one uses lens systems made of different kinds of glass.

8.46 Example (**Achromatic Lens**)
An *achromatic lens* or, more precisely, *achromatic doublet*, is a system of two lenses, often consisting of a collecting lens made of crown glass and a diverging lens made of flint glass, that are often glued together, see figure. If the achromatic doublet is to have a focal length $f(\omega_0)$ at frequency ω_0, it needs to satisfy



$$\frac{1}{f(\omega_0)} = \frac{1}{f_F(\omega_0)} + \frac{1}{f_K(\omega_0)} . \tag{8.7.6}$$

If moreover $f'(\omega_0) = 0$ is required, then by (8.7.4), the condition

$$\frac{r_F(\omega_0)}{f_F(\omega_0)} + \frac{r_K(\omega_0)}{f_K(\omega_0)} = 0 \quad \text{with} \quad r_F = \frac{n'_F(\omega_0)}{n_F(\omega_0) - 1} \text{ and } r_K := \frac{n'_K(\omega_0)}{n_K(\omega_0) - 1} \tag{8.7.7}$$

must be verified. Equations (8.7.6) and (8.7.7) are linear in the reciprocals of the unknown focal lengths $f_F(\omega_0)$ and $f_K(\omega_0)$. They are solvable, given $f(\omega_0)$, because the coefficients r_F and r_K for flint glass and crown glass are unequal. The solution is[11]

$$\frac{1}{f_F} = \frac{r_K}{r_K - r_F} \frac{1}{f(\omega_0)} \quad , \quad \frac{1}{f_K} = \frac{r_F}{r_F - r_K} \frac{1}{f(\omega_0)} .$$

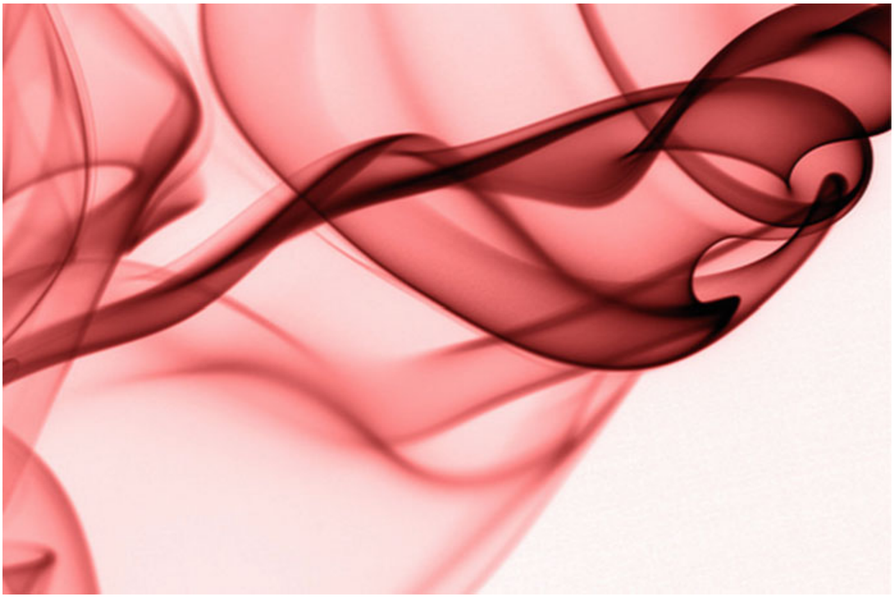
◊

8.47 Remark (**Linear Optics and Symplectic Group**)
We have observed in this chapter that geometrical optics in its linear approximation leads to calculations in the group $\mathrm{Sp}(4, \mathbb{R})$.

Indeed, one can show (see Guillemin and Sternberg [GS1], Chapter 4) that linear optics is equivalent to the theory of this symplectic group in the sense that every element of this group can be written as a finite product of matrices of the forms (8.7.1) and (8.7.3). ◊

[11] In practice, one frequently corrects for the colors blue and red (see figure) and therefore uses quantities called *Abbe numbers* of the different kinds of glass, rather than their refractive indices.

Chapter 9
Ergodic Theory



Smoke. Picture: courtesy of Rick Hanley.

In ergodic theory, one investigates statistical properties of dynamical systems. This is often possible even if the dynamics is difficult to calculate, namely in the case of chaotic motion (and particularly in this case).

9.1 Measure Preserving Dynamical Systems

> *"The studies about the foundations of geometry suggest to us the problem of treating, according to this paradigm, those disciplines of physics in which mathematics is already today playing a prominent role; primarily these are probability and mechanics."*
> DAVID HILBERT, 6th problem

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_9

After the topological and geometric properties of dynamical systems, we will now look at aspects of measure theory and probability. This point of view is particularly important when we describe the long term behavior of unstable ('chaotic') systems.

In the simplest case, the measure under consideration will be the Lebesgue measure λ^d on the phase space $\mathbb{R}^d$. However, as we will also consider other measures, we start with some basic notions of measure and probability theory.

The above-quoted sixth of Hilbert's 23 problems presented in his speech at the International Congress of Mathematicians in Paris, in 1900, shows that at that time, the theory of probability had not been put completely on a mathematical foundation yet (and was partly attributed to physics, as an applied science, due to the statistical mechanics propagated by Boltzmann, among others). This axiomatization happened essentially by Andrey Kolmogorov in his 1933 textbook *Foundations of the Theory of Probability*.

9.1 Definition A **measurable space** $(M, \mathcal{M})$ *is a nonempty set* M *with a family* $\mathcal{M}$ *of subsets of* M (the **measurable sets**), *satisfying*:

- $M \in \mathcal{M}$.
- If $A_n \in \mathcal{M}$ $(n \in \mathbb{N})$, then also $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$.
- If $A \in \mathcal{M}$, then also $A^c := M \setminus A \in \mathcal{M}$.

$\mathcal{M}$ is then called a σ-**algebra** on M.

Compare this with the definition A.1 of topological spaces.

9.2 Examples (Measurable spaces)

1. $\{\emptyset, M\}$ is the smallest σ-algebra on M.
2. The power set $2^M = \mathcal{P}(M)$ of M is the largest σ-algebra on M.
3. If $M \neq \emptyset$ and $\mathcal{N}$ a subset of the power set 2^M, then the system

$$\sigma(\mathcal{N}) := \bigcap_{\sigma-\text{algebra } \mathcal{A}\subseteq 2^M:\ \mathcal{N}\subseteq\mathcal{A}} \mathcal{A} \tag{9.1.1}$$

   is a σ-algebra on M with $\mathcal{N} \subseteq \sigma(\mathcal{N})$. It is the smallest such σ-algebra and is called the *σ-algebra generated by $\mathcal{N}$*.
4. If $(M, \mathcal{O})$ is a topological space, then $\sigma(\mathcal{O})$ is called the σ-algebra of the *Borel sets*. It contains in particular all open and all closed subsets of M.
   This is the σ-algebra used most frequently in classical mechanics on the phase space M (which, as a manifold, is in particular a topological space). ◊

9.3 Definition

- *A* **measure** *on a measurable space* $(M, \mathcal{M})$ *is a mapping*[1] $\mu : \mathcal{M} \to [0, \infty]$ (*that is not constant* ∞), *which is* σ-**additive** (*also called* **countably additive**), *i.e.,*

[1] $[0, \infty]$ denotes the set $[0, +\infty) \cup \{+\infty\}$.

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n)\,, \tag{9.1.2}$$

for disjoint ($A_m \cap A_n = \emptyset$ *for* $m \neq n \in \mathbb{N}$) *sets* $A_n \in \mathcal{M}$.

- μ *is called a* **probability measure** *if moreover* $\mu(M) = 1$.
- A **measure space** $(M, \mathcal{M}, \mu)$ *is a measurable space*$(M, \mathcal{M})$ *together with a measure*$\mu : \mathcal{M} \to [0, \infty]$.
- *If* μ *is a probability measure*, *then* $(M, \mathcal{M}, \mu)$ *is called a* **probability space**.

9.4 Examples (Measures)

1. $(\mathbb{R}^d, +)$ and $(\mathbb{Z}^d, +)$ are examples of (locally compact) abelian **topological groups** $(M, +)$. For such groups, there exists a translation invariant[2](regular) measure μ on a certain σ-algebra $\mathcal{M}$ containing the Borel sets of M that is positive on nonempty open subsets.
   This measure is unique up to multiplication by a constant and is called *Haar measure*. Its construction can be found in ELSTRODT [El], Chapter VIII, §3.
   For instance, in the case of the group $(\mathbb{R}^d, +)$, we obtain $\mu = c\lambda^d$ for the Lebesgue measure λ^d with a normalization constant $c := \mu\big([0, 1]^d\big) > 0$.
2. If h is a **regular value** of a smooth function $H : \mathbb{R}^d \to \mathbb{R}$, and if the level set $M := \{x \in \mathbb{R}^d \mid H(x) = h\}$ is compact (and $M \neq \emptyset$), then

$$\lambda_h(B) := \lim_{\varepsilon \searrow 0} \frac{\lambda^d\Big(\hat{B} \cap H^{-1}\big((h-\varepsilon, h+\varepsilon)\big)\Big)}{2\varepsilon}$$

   defines a measure on M. In this definition, we have assigned to any Borel set $B \in \mathcal{M}$ the set $\hat{B} \subseteq \mathbb{R}^d$ that is intuitively a 'thickening' of B; more precisely it is the union of all segments of some small length 2δ that are centered at some $b \in B$ and orthogonal to M. If H is a Hamilton function on phase space, the measure λ_h on M is called *Liouville measure*.[3] It can be proved that the limit indeed exists. But despite the notation, it does depend on the function H, not merely the value h. For example, for $H : \mathbb{R}^{d+1} \to \mathbb{R}$, $H(x) := \|x\|^2$, we obtain a rotationally symmetric measure on the sphere $S^d = H^{-1}(1)$. ◊

[2]**Definition:** *Translation invariance* of μ means $\mu(A + m) = \mu(A)$ for all $A \in \mathcal{M}$ and $m \in M$.

[3]See [AM, Theorem 3.4.12] for the construction of a corresponding volume form.

9.5 Definition

- *A mapping* $T : M_1 \to M_2$ *between measure spaces* $(M_i, \mathcal{M}_i, \mu_i)$ *is called* **measurable** *if*
$$T^{-1}(A_2) \in \mathcal{M}_1 \qquad (A_2 \in \mathcal{M}_2).$$
- *A measurable mapping* $T : M_1 \to M_2$ *is called* **measure preserving** if
$$\mu_1\left(T^{-1}(A_2)\right) = \mu_2(A_2) \qquad (A_2 \in \mathcal{M}_2).$$
- If $\Phi : G \times M \to M$ *is a dynamical system (with group* $G = \mathbb{R}$ *or* $\mathbb{Z}$*), and* $(M, \mathcal{M}, \mu)$ *is a measure space, then* $(M, \mathcal{M}, \mu, \Phi)$ *is called a* **measure preserving dynamical system** *if* Φ *is measurable*[4] *and the mappings* $\Phi_t : M \to M$ $(t \in G)$ *are measure preserving.*

9.6 Exercise (Invariant Measure)
Show that the piecewise continuous **Gauss map**[5]

$$h : [0, 1) \to [0, 1) \quad , \quad h(0) := 0 \quad , \quad h(x) := 1/x - \lfloor 1/x \rfloor \text{ for } x > 0$$

preserves the probability measure μ on [0, 1] given by $\mu(A) := \frac{1}{\log 2} \int_A \frac{1}{1+x}\, \mathrm{d}x$. ◊

9.7 Remark (Existence of Invariant Measures)
In most cases, one knows that a given dynamical system has an invariant measure. For example, the theorem by Bogoliubov and Krylov guarantees for continuous mappings $T : M \to M$ of a compact metric space M even the existence of a T-invariant probability measure.[6] ◊

In classical mechanics, there is always a distinguished invariant measure:

9.8 Theorem *Let* $H \in C^2(M, \mathbb{R})$ *on the phase space* $M := \mathbb{R}^n_p \times \mathbb{R}^n_q$ *be a function generating (according to Definition 6.6) a Hamiltonian flow* Φ *on* M*. Then the mappings* $\Phi_t : M \to M$ $(t \in \mathbb{R})$ *preserve the Lebesgue measure* λ^{2d}*.*

Proof For the case of a linear flow, this claim was already proved as a corollary to Theorem 6.11. Analogously, the claim now follows using Theorem 4.18 (Wronskian) from the form $X_H = \mathbb{J}\,\mathrm{D}H$ of the Hamiltonian vector field of H. Namely, the trace of the linearized vector field is 0:

$$\mathrm{tr}(\mathrm{D}X_H) = \mathrm{tr}\big(\mathbb{J}\,\mathrm{D}^2 H\big) = \mathrm{tr}\big((\mathbb{J}\,\mathrm{D}^2 H)^\top\big) = \mathrm{tr}\big(-\mathbb{J}\,\mathrm{D}^2 H\big) = -\mathrm{tr}(\mathrm{D}X_H). \qquad \square$$

[4] with $G \times M$ carrying the product σ-algebra of the Borel σ-algebra on G and $\mathcal{M}$.

[5] See also 433.

[6] It is constructed as a cluster point of a sequence of measures that come closer and closer to being invariant, see WALTERS [Wa2], *Corollary* 6.9.1.

9.9 Remark (Hamiltonian Flows are Measure Preserving)
Compare with Remark 10.14, concerning Hamiltonian flows on symplectic manifolds. Analogously, for regular values E of H, the flow on the energy surface $H^{-1}(E)$ obtained by restriction preserves the Liouville measure λ_E. ◊

9.10 Exercise (Phase Space Volume) Suppose we are given the Hamiltonian

$$H : M \to \mathbb{R}\,,\; H(p,q) := \tfrac{1}{2}\|p\|^2 + W(\|q\|) \quad \text{with potential } W(r) := -r^{-\alpha}$$

on the phase space $M := \mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\})$. For which values of the parameter $\alpha \in \mathbb{R}$ is the volume $V(E) := \lambda^{2n}\left(\left\{(p,q) \in M \,\middle|\, H(p,q) \le E\right\}\right)$ of the part of the phase space below energy E finite for all $E < 0$, for which is it infinite? ◊

9.2 Ergodic Dynamical Systems

Many dynamical systems have the property that initially nearby trajectories diverge rapidly, thus making long-term predictions on the basis of finite precision knowledge of the initial data impossible.

Nevertheless, it is possible in these cases as well to make statements about long-term behavior; however, these statements will be of a statistical nature.

This application of probability theoretic notions to dynamical systems is called *ergodic theory*. So we investigate the properties of a measure preserving dynamical system $(M, \mathcal{M}, \mu, \Phi)$, where we assume that μ is a probability measure.[7]

An important question is how many Φ-invariant measurable sets there are, in other words, how large the set

$$\mathcal{I} := \{A \in \mathcal{M} \mid \forall t \in G : \Phi_t(A) = A\} \tag{9.2.1}$$

actually is. Trivially, it is always true that $\{\emptyset, M\} \subset \mathcal{I}$, and $\mathcal{I} \subset \mathcal{M}$ is a σ-algebra on M. Roughly speaking, $\mathcal{I}$ is the smaller the more the flow Φ mixes up the phase space M.

9.11 Definition
The measure preserving dynamical system is called **ergodic**[8] *if*

$$\mu(A) \in \{0, 1\} \qquad (A \in \mathcal{I}).$$

9.12 Examples (Ergodicity of Circle Rotations)
Let M be the circle $S^1 \subset \mathbb{C}$. As this phase space is also a compact abelian group, the appropriately normed Haar measure[9] μ on M is a probability measure.

---

[7] There does also exist an ergodic theory for measures that are not finite, see AARONSON [Aa].
[8] To preempt any misunderstandings: Ergodic theory also studies systems that are not ergodic.
[9] See Example 9.4 on 193.

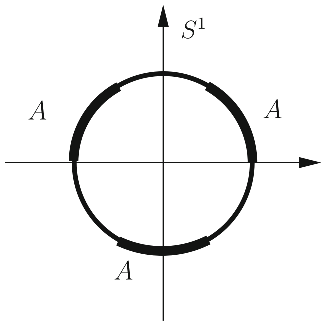
1. The group action of $\mathbb{R}$ given by $\Phi : \mathbb{R} \times M \to M$, $\Phi(t, m) := \exp(2\pi\imath\alpha t)\, m$ is measure preserving and is ergodic for $\alpha \in \mathbb{R} \setminus \{0\}$, because if any $m \in S^1$ belongs to a Φ-invariant set $A \subseteq S^1$, then $A = S^1$.
2. For $\alpha \in \mathbb{Q}$, the group action of $\mathbb{Z}$,

$$\Phi : \mathbb{Z} \times M \to M \ , \ \Phi(t, m) := \exp(2\pi\imath\alpha t)\, m$$

is not ergodic: namely if $\alpha = p/q$ with $q \in \mathbb{N}$ and $p \in \mathbb{Z}$, then the set

$$A := \bigcup_{k=0}^{q-1} \exp\left(2\pi\imath \left[\frac{k}{q}, \frac{k+\frac{1}{2}}{q}\right]\right) \subset M$$

is Φ-invariant, but $\mu(A) = \frac{1}{2}$, see figure on the right. ♢



Φ-invariant set A for $\alpha = p/q$ with $q = 3$

In the last example, for $\alpha \in \mathbb{Q}$ all orbits are periodic, whereas for $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ no orbit is periodic. In that case, will this discrete dynamical system then also be ergodic?

This question isn't answered quite easily; certainly there are more invariant sets than M and $\emptyset$ (for example the orbit $\{\exp(2\pi\imath\alpha t)x \mid t \in \mathbb{Z}\}$ of a single point $x \in M$; this orbit is countable and thus does not coincide with M).

In this case, it is useful to consider how the dynamical system operates on phase space functions, specifically the square integrable functions on M.

9.13 Lemma *If $(M, \mathcal{M}, \mu, \Phi)$ is a measure preserving dynamical system, then the linear endomorphisms*

$$\hat{\Phi}_t : L^2(M, \mu) \to L^2(M, \mu) \quad , \quad \hat{\Phi}_t f := f \circ \Phi_t \qquad (t \in G)$$

are unitary, i.e., $\hat{\Phi}_t$ is surjective and $\left\langle \hat{\Phi}_t f, \hat{\Phi}_t g \right\rangle = \langle f, g \rangle \quad \big(f, g \in L^2(M, \mu)\big)$.

Proof

- By the Φ_t-invariance of the measure μ, it follows with $y = \Phi_t(x)$ that

$$\begin{aligned} \left\langle \hat{\Phi}_t f, \hat{\Phi}_t g \right\rangle &= \int_M f \circ \Phi_t(x)\, \overline{g} \circ \Phi_t(x)\, \mathrm{d}\mu(x) = \int_M f(y)\, \overline{g}(y)\, \mathrm{d}\mu\big(\Phi_{-t}(y)\big) \\ &= \int_M f(y)\, \overline{g}(y)\, \mathrm{d}\mu(y) = \langle f, g \rangle \,. \end{aligned}$$

- $\hat{\Phi}_t$ is surjective, since the inverse mapping exists: $(\hat{\Phi}_t)^{-1} = \hat{\Phi}_{-t}$. □

9.14 Theorem (Koopman)
The group action Φ is ergodic if and only if all $\hat{\Phi}$-invariant functions

$$f \in L^2(M,\mu) \quad , \quad \hat{\Phi}_t f = f \quad (t \in G)$$

are constant μ-almost everywhere.

Proof We will identify the square integrable functions $f : M \to \mathbb{C}$ with their equivalence classes $[f] \in L^2(M,\mu)$.

- For a $\hat{\Phi}$-invariant function $f \in L^2(M,\mu)$, the functions Re (f) and Im (f) will be $\hat{\Phi}$-invariant as well, and vice versa. So we may assume that f is real valued. For all $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, the measurable subsets

$$A_{n,k} := f^{-1}\big([k2^{-n}, (k+1)2^{-n})\big)$$

  of the phase space are Φ-invariant ($A_{n,k} \in \mathcal{I}$). For all $n \in \mathbb{N}$, they also form a partition of M, i.e.,

$$A_{n,k_1} \cap A_{n,k_2} = \emptyset \text{ for } k_1 \neq k_2 \in \mathbb{Z} \quad \text{, and} \quad \bigcup_{k\in\mathbb{Z}} A_{n,k} = M\,. \qquad (9.2.2)$$

  If Φ is ergodic, then $\mu(A_{n,k}) \in \{0,1\}$; consequently by (9.2.2), there exists a unique index k_n with $\mu(A_{n,k_n}) = 1$. Now

$$A_{n,k} = A_{n+1,2k} \,\dot{\cup}\, A_{n+1,2k+1}\,,$$

  and therefore the sequence $(k_n 2^{-n})_{n\in\mathbb{N}}$ converges to a real number z, which is taken on by f μ-almost everywhere (i.e., $\mu(\{x \in M \mid f(x) \neq z\}) = 0$).
- Conversely, if all $\hat{\Phi}$-invariant functions $f \in L^2(M,\mu)$ are constant μ-almost everywhere, then this applies in particular to the characteristic functions $\mathbb{1}_A$ of invariant sets $A \in \mathcal{I}$. As these take on only the values 0 and 1, it follows

$$\mu(A) = \int_M \mathbb{1}_A \, d\mu \in \{0,1\} \qquad (A \in \mathcal{I}),$$

  hence ergodicity. □

9.15 Examples (Ergodicity of Discrete Circle Rotation, Continued)
We return to Example 9.12.2 and assume that the circle S^1 is rotated by integer multiples of the angle $2\pi\alpha$, with $\alpha \in \mathbb{R}\setminus\mathbb{Q}$. Now if $f \in L^2(M,\mu)$ is invariant under $\hat{\Phi}$ and we let

$$f = \sum_{k\in\mathbb{Z}} c_k e_k \quad \text{with} \quad c_k := \langle f, e_k\rangle$$

be the Fourier expansion of f with the orthonormal basis consisting of the characters

$$e_k : S^1 \to S^1 \quad , \quad e_k(m) := m^k \qquad (k \in \mathbb{Z}),$$

then

$$\hat{\Phi}_t(f) \;=\; \sum_{k\in\mathbb{Z}} c_k \hat{\Phi}_t(e_k) \;=\; \sum_{k\in\mathbb{Z}} c_k \exp(2\pi\imath tk\alpha)\, e_k\,,$$

and for time $t = 1$ it follows: $\sum_{k\in\mathbb{Z}} c_k\big(1 - \exp(2\pi\imath k\alpha)\big)\, e_k = f - \hat{\Phi}_1(f) = 0$.

Since α is irrational, the parenthesis only vanishes when $k = 0$, and therefore the basis property of the functions e_k implies that: $c_k = 0 \quad (k \in \mathbb{Z} \setminus \{0\})$.

Therefore, f is constant μ-almost everywhere. Thus Φ is ergodic. ◊

9.3 Mixing Dynamical Systems

"Ich bin eigentlich nur Physiker aus Ordnungsliebe geworden (er stellt die Stehlampe auf). Um die scheinbare Unordnung in der Natur auf eine höhere Ordnung zurückzuführen."
NEWTON, in *The Physicists* by Friedrich Dürrenmatt[10]

As shown in the last example, very regular and well predictable dynamical systems can be ergodic. In contrast, the dynamics of mixing systems is more complicated.

We again begin with a measurable dynamical system $(M, \mathcal{M}, \mu, \Phi)$ with a probability measure μ.

9.16 Definition *The dynamical system is called* **mixing**, *if*

$$\lim_{|t|\to\infty} \mu\big(\Phi_t(A) \cap B\big) = \mu(A)\,\mu(B) \qquad (A, B \in \mathcal{M}). \tag{9.3.1}$$

In a measure theoretic sense, for large times t, the set $\Phi_t(A)$ is equally distributed in M, somewhat like milk gets distributed in coffee by stirring.

9.17 Lemma *Mixing dynamical systems are ergodic.*

Proof If $A \in \mathcal{I}$ (see (9.2.1)), then $\Phi_t(A) = A$ for all times t, hence by the mixing property, $\mu(A \cap B) = \lim_{|t|\to\infty} \mu(\Phi_t(A) \cap B) = \mu(A)\,\mu(B)$.
For $B := A$, this yields the equation $\mu(A) = \mu(A)^2$, hence $\mu(A) \in \{0, 1\}$. □

9.18 Example The example of rotations of the circle (Example 9.12.1 and 9.15) shows that conversely, not all ergodic dynamical systems are mixing, because in this example the limit from (9.3.1) does not exist, see Figure 9.3.1. ◊

[10] Translation: "I can't stand disorder. Actually, I became a physicist only because of my love of order. (He rights the floor lamp). To prove that the apparent disorder of nature is founded in a higher order."

However, it is true for ergodic systems that the Cesàro-mean[11] of the function on the left hand side of (9.3.1) converges to the right hand side.

9.19 Definition

- *For a measure preserving dynamical system* $(M, \mathcal{M}, \mu, \Phi)$ *with probability measure* μ*, the group of times* $G = \mathbb{R}$ or $G = \mathbb{Z}$*, and* $f, g \in L^2(M, \mu)$*, the function*

$$C_{f,g} : G \to \mathbb{C} \quad , \quad C_{f,g}(t) := \left\langle \hat{\Phi}_t f, g \right\rangle - \langle f, 1\!\!1_M \rangle \langle 1\!\!1_M, g \rangle$$

is called the **correlation function** of f and g.
- $C_f := C_{f,f}$ *is called the* **autocorrelation function** of f.

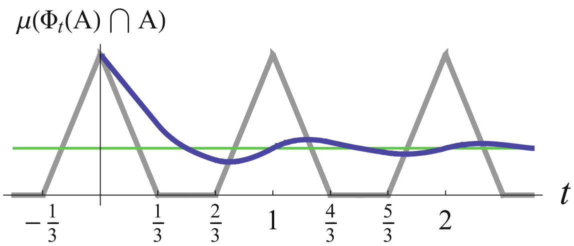


Figure 9.3.1 The function $t \mapsto \mu\big(\Phi_t(A) \cap A\big)$, its mean, and the Cesàro-means, for the circle rotation $\Phi_t(x) = \exp(2\pi\imath t)\, x$ and the circle segment $A := \{\exp(2\pi\imath z) \mid 0 \le z \le 1/3\}$

In particular, the correlation function of characteristic functions is of the form

$$C_{1\!\!1_A}, 1\!\!1_B(t) = \mu\big(\Phi_t(A) \cap B\big) - \mu(A)\,\mu(B) \qquad (A, B \in \mathcal{M}). \tag{9.3.2}$$

$C_{f,g}$ is linear in the function f and conjugate linear in the function g.
Moreover, $C_{\tilde{f},\tilde{g}} = C_{f,g}$ if $\tilde{f} - f$ and $\tilde{g} - g$ are constant functions.

In analogy to Theorem 9.14, we have

9.20 Theorem *For a measure preserving dynamical system* $(M, \mathcal{M}, \mu, \Phi)$ *with probability measure* μ*, the following statements are equivalent:*

1. *The system is mixing.*
2. *The correlation function of* $f, g \in L^2(M, \mu)$ *satisfies* $\lim_{|t|\to\infty} C_{f,g}(t) = 0$.
3. *The autocorrelation function of* $f \in L^2(M, \mu)$ *satisfies* $\lim_{|t|\to\infty} C_f(t) = 0$.

Proof

- For $f := 1\!\!1_A$ and $g := 1\!\!1_B$, one obtains from (9.3.2) the implication #2 ⇒ #1.

---

[11] **Definition** The *Cesàro-mean* of a sequence $(a_k)_{k\in\mathbb{N}}$ is the sequence $(c_k)_{k\in\mathbb{N}}$, $c_k := \frac{1}{k}\sum_{\ell=1}^{k} a_\ell$. The *Cesàro-mean* of an integrable function $a : \mathbb{R}^+ \to \mathbb{C}$ is the function $c : \mathbb{R}^+ \to \mathbb{C}$ with $c(x) := \frac{1}{x}\int_0^x a(y)\,\mathrm{d}y$.

- #1 $\Rightarrow$ #3 follows by approximation of $f \in L^2(M,\mu)$ by *simple functions*, i.e., functions of the form $f_n := \sum_{k=1}^n c_{n,k}\,1\!\!1_{A_{n,k}}$, with coefficients $c_{n,k} \in \mathbb{C}$ and measurable $A_{n,k} \subseteq M$. For $\varepsilon > 0$, let $N(\varepsilon)$ be chosen such that

$$\|f - f_n\|_2 < \varepsilon \qquad \big(n \geq N(\varepsilon)\big).$$

It suffices to show property 3 for functions f with mean $\langle f, 1\!\!1_M\rangle = 0$, and these latter can be approximated by simple functions f_n with mean $\langle f_n, 1\!\!1_M\rangle = 0$. From the Cauchy-Schwarz inequality and the unitary property of $\hat{\Phi}_t$ (Lemma 9.13), it follows that

$$\begin{aligned} |C_f(t) - C_{f_n}(t)| &= \left|\left\langle \hat{\Phi}_t(f - f_n), f\right\rangle + \left\langle \hat{\Phi}_t f_n, f - f_n\right\rangle\right| \\ &\leq \|\hat{\Phi}_t(f - f_n)\|_2\|f\|_2 + \|\hat{\Phi}_t f_n\|_2\|f - f_n\|_2 \\ &= \|f - f_n\|_2(\|f\|_2 + \|f_n\|_2) \leq \varepsilon(\|f\|_2 + \|f\|_2 + \varepsilon)\,, \end{aligned} \tag{9.3.3}$$

and therefore $\lim_{n\to\infty} C_{f_n} = C_f$ uniformly in time. On the other hand, by hypothesis 1,

$$C_{f_n}(t) = \sum_{k,\ell=1}^{n} c_{n,k}\overline{c_{n,\ell}}\,\Big(\mu\big(\Phi_t(A_{n,k}) \cap A_{n,\ell}\big) - \mu(A_{n,k})\,\mu(A_{n,\ell})\Big)$$

converges to 0 for all $n \in \mathbb{N}$ in the limit $|t| \to \infty$.
- #3 $\Rightarrow$ #2 follows from a polarization identity. If we assume again, without loss of generality, that $f, g \in L^2(M,\mu)$ have mean zero, the same is true for $h_k := \frac{1}{2}(f + i^k g)$, and

$$\begin{aligned} \sum_{k=1}^{4} i^k C_{h_k}(t) &= \tfrac{1}{4}\sum_{k=1}^{4}\Big[\left\langle \hat{\Phi}_t f, g\right\rangle + (-1)^k\left\langle \hat{\Phi}_t g, f\right\rangle + i^k\left\langle \hat{\Phi}_t f, f\right\rangle + i^k\left\langle \hat{\Phi}_t g, g\right\rangle\Big] \\ &= \left\langle \hat{\Phi}_t f, g\right\rangle = C_{f,g}(t)\,. \end{aligned} \tag{9.3.4}$$

As by hypothesis, the left side of (9.3.4) converges to 0 in the limit $|t| \to \infty$, it is also true that $\lim_{|t|\to\infty} C_{f,g}(t) = 0$. □

9.21 Examples (Torus Automorphisms)
The set of all invertible 2×2-matrices with integer entries is a multiplicative group

$$\mathrm{GL}(2,\mathbb{Z}) := \{T \in \mathrm{Mat}(2,\mathbb{Z}) \mid |\det(T)| = 1\}\,,$$

the *general linear group* of degree 2 over $\mathbb{Z}$.

For each $T \in \mathrm{GL}(2,\mathbb{Z})$, we obtain thus a dynamical system

$$\tilde{\Phi} : \mathbb{Z} \times \mathbb{R}^2 \to \mathbb{R}^2 \quad , \quad \tilde{\Phi}(n,x) := T^n x\,,$$

which maps the lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$ to itself.

Now $\mathbb{Z}$ is a subgroup of the abelian group $\mathbb{R}$, so the set $\mathbb{R}/\mathbb{Z}$ of cosets is also an abelian group (by Theorem E.7), and we can choose the numbers from $[0, 1)$ as representatives. Then addition of real numbers becomes addition modulo 1.

The mapping $x \mapsto \exp(2\pi\imath x)$ maps $\mathbb{R}/\mathbb{Z}$ isomorphically onto the multiplicative group $S^1 = \{c \in \mathbb{C} \mid |c| = 1\}$, the circle. Componentwise addition modulo 1 produces the isomorphism of the 2-torus $\mathbb{T}^2 = S^1 \times S^1$ as introduced in Example 8.32 (double pendulum) with the factor group

$$\hat{\mathbb{T}}^2 := \mathbb{R}^2/\mathbb{Z}^2 \cong (\mathbb{R}/\mathbb{Z})^2 \quad \text{and projection } \pi : \mathbb{R}^2 \to \hat{\mathbb{T}}^2 \,,\ \binom{x_1}{x_2} \mapsto \begin{pmatrix} x_1 \ (\mathrm{mod}\, 1) \\ x_2 \ (\mathrm{mod}\, 1) \end{pmatrix}.$$

As the mappings $\tilde{\Phi}_t$ are linear and $\tilde{\Phi}_t(\mathbb{Z}^2) = \mathbb{Z}^2$, one obtains for all times $t \in \mathbb{Z}$ that

$$\tilde{\Phi}_t(x + \ell) \equiv \tilde{\Phi}_t(x) \pmod 1 \qquad \big(x \in \mathbb{R}^2,\ \ell \in \mathbb{Z}^2\big).$$

Thus, by projecting to the factor group, we get the dynamical system

$$\Phi_t : \hat{\mathbb{T}}^2 \to \hat{\mathbb{T}}^2 \quad , \quad \Phi_t\big(\pi(x)\big) = \pi\big(\tilde{\Phi}_t(x)\big) \qquad \big(x \in \mathbb{R}^2\big).$$

Φ_1 is an automorphism of the group $\hat{\mathbb{T}}^2$ and is therefore called a *torus automorphism*.[12] If moreover, $\det(T) = 1$, then one calls T and the torus automorphism

- *hyperbolic* if $|\mathrm{tr}(T)| > 2$, i.e., if T has real eigenvalues λ_i with $|\lambda_1| > 1 > |\lambda_2|$. The mapping T will then stretch by a factor λ_1 along the direction of the first eigenspace, and contract by a factor $\lambda_2 = 1/\lambda_1$ along the direction of the second eigenspace.
- *parabolic* if $|\mathrm{tr}(T)| = 2$, i.e., the eigenvalue is equal 1 or -1.
- *elliptic* if $|\mathrm{tr}(T)| < 2$, in which case the linear mapping given by T is conjugate to a rotation of the plane by an angle of $\pm\pi/3$, $\pm\pi/2$ or $\pm 2\pi/3$.

In Figures 9.3.2 and 9.3.3 we see the operation of Φ_t for a hyperbolic or a parabolic matrix T, respectively, on a subset of the 2-torus $\hat{\mathbb{T}}^2$. ◊

9.22 Theorem *Hyperbolic torus automorphisms are mixing.*

Proof The Hilbert space $L^2(\hat{\mathbb{T}}^2)$ of (equivalence classes of) square integrable functions $f, g : \hat{\mathbb{T}}^2 \to \mathbb{C}$ with the scalar product $\langle f, g\rangle := \int_{\hat{\mathbb{T}}^2} f(x)\overline{g}(x)\, dx$ has the orthonormal basis of characters $(e_k)_{k\in\mathbb{Z}^2}$ with

$$e_k(x) := \exp(2\pi\imath\, \langle k, x\rangle) \qquad \big(x \in \hat{\mathbb{T}}^2\big). \tag{9.3.5}$$

These characters satisfy

$$e_k\big(\Phi_n(x)\big) = \exp\big(2\pi\imath\, \langle k, \Phi_n(x)\rangle\big) = \exp\left(2\pi\imath \left\langle \tilde{\Phi}_n^\top(k), x\right\rangle\right) = e_{\tilde{\Phi}_n^\top(k)}(x)\,,$$

[12] As a matter of fact, every continuous automorphism of the group $\hat{\mathbb{T}}^2$ has this form (Walters [Wa2], §0.8).

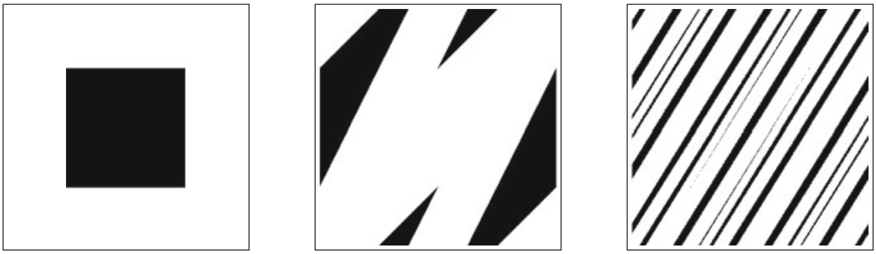


Figure 9.3.2 Operation of the hyperbolic torus automorphism for the matrix $T := \binom{2\,1}{1\,1}$. Left: a subset $A \subset \mathbb{T}^2$. Center: $\Phi_1(A)$, Right: $\Phi_3(A)$

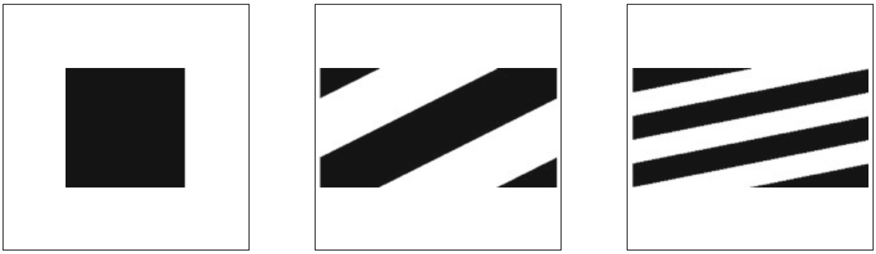


Figure 9.3.3 Operation of the parabolic torus automorphism for the matrix $T := \binom{1\,1}{0\,1}$. Left: a subset $A \subset \mathbb{T}^2$. Center: $\Phi_2(A)$, Right: $\Phi_5(A)$

because if T is in $\mathrm{GL}(2, \mathbb{Z})$, then so is the transpose $T^\top$.

Since T is hyperbolic, there does not exist a lattice vector $k \in \mathbb{Z}^2 \setminus \{0\}$ that would be mapped to itself for any time $t \in \mathbb{Z} \setminus \{0\}$; so $\tilde{\Phi}_t^\top(k) \neq k$.

Let us therefore consider, for $r > 0$ and arbitrarily large, the finite set $\mathbb{Z}_r^2 := \{k \in \mathbb{Z}^2 \mid |k| \leq r\}$; then there exists a time $t_0(r)$, by which all $k \in \mathbb{Z}_r^2 \setminus \{0\}$ should have left this set, i.e.,

$$\tilde{\Phi}_t^\top(k) \notin \mathbb{Z}_r^2 \qquad \big(|t| > t_0(r)\big).$$

The linear mappings

$$\hat{\Phi}_t : L^2(\hat{\mathbb{T}}^2) \to L^2(\hat{\mathbb{T}}^2) \quad , \quad \hat{\Phi}_t f = f \circ \Phi_t$$

are unitary by Lemma 9.13. We show, according to Theorem 9.20, that

$$\lim_{|t|\to\infty} C_f(t) = 0 \qquad \big(f \in L^2(\hat{\mathbb{T}}^2)\big). \tag{9.3.6}$$

We again may assume without loss of generality that f averages to 0, so that in the Fourier series

$$f = \sum_{k\in\mathbb{Z}^2} c_k\, e_k \quad \text{with} \quad c_k := \langle f, e_k\rangle \in \mathbb{C}$$

(and $\sum_{k\in\mathbb{Z}^2} |c_k|^2 < \infty$) the coefficient $c_0 = 0$. We truncate this series, letting

$$f_r := \sum_{n\in\mathbb{Z}_r^2} c_n\, e_n\,.$$

Now for times $t \in \mathbb{Z}$ with $|t| > t_0(r)$, one has

$$\left\langle \hat{\Phi}_t f_r,\, f_r \right\rangle = \sum_{k,\ell\in\mathbb{Z}_r^2} c_k\, \overline{c}_\ell \cdot \left\langle \hat{\Phi}_t e_k,\, e_\ell \right\rangle = |c_0|^2 = 0\,,$$

because $\hat{\Phi}_t e_k = e_{\tilde{\Phi}_t^\top(k)}$ is then orthonormal on the e_ℓ, $\ell \in \mathbb{Z}_r^2$ for lattice points $k \in \mathbb{Z}_r^2 \setminus \{0\}$. On the other hand, the mappings $\hat{\Phi}_t$ are unitary, so with the triangle inequality and the Cauchy-Schwarz inequality, it follows, analog to (9.3.3), that

$$\left| \left\langle \hat{\Phi}_t f,\, f \right\rangle - \left\langle \hat{\Phi}_t f_r,\, f_r \right\rangle \right| \le \varepsilon(2\|f\|_2 + \varepsilon)\,,$$

provided $r \equiv r(\varepsilon)$ is chosen so large that $\|f - f_r\|_2 < \varepsilon$. For $|t| > t_0(r(\varepsilon))$, this implies the inequality $|C_f(t)| \le \varepsilon(2\|f\|_2 + \varepsilon)$, hence (9.3.6). □

9.23 Exercise (Decay of Correlation) By Theorem 9.22, a hyperbolic torus automorphism $T : \mathbb{T}^2 \to \mathbb{T}^2$, $Tx = \hat{T}x \pmod 1$ with $\hat{T} \in \mathrm{GL}(2,\mathbb{Z})$ and $\left|\mathrm{tr}(\hat{T})\right| > 2$ is mixing, i.e., with $U : L^2(\mathbb{T}^2) \to L^2(\mathbb{T}^2)$ and $Ug := g \circ T$, one has

$$\lim_{n\to\infty} \langle f, U^n g\rangle = \langle f, \mathbb{1}\rangle\, \langle \mathbb{1}, g\rangle \qquad \big(f, g \in L^2(\mathbb{T}^2)\big).$$

In this exercise, we estimate the rate of convergence for $f, g \in C^1(\mathbb{T}^2)$.

To this end, we use that the Fourier coefficients $f_k = \langle f, e_k\rangle \in \mathbb{C}$, [4] $k \in \mathbb{Z}^2$ satisfy more than merely the Parseval equality $\sum_{k\in\mathbb{Z}^2} |f_k|^2 = \langle f, f\rangle$, which follows from the orthonormality of the characters $e_k : \mathbb{T}^2 \to S^1$, $e_k(x) = \exp(2\pi\imath\, \langle k, x\rangle)$; if $f \in C^1(\mathbb{T}^2)$, it even follows that

$$\sum_{k\in\mathbb{Z}^2} \|k\|^2\, |f_k|^2 \;=\; \frac{\|\nabla f\|_2^2}{4\pi^2} \;<\; \infty\,. \tag{9.3.7}$$

So the Fourier coefficients have to decay comparatively rapidly, because f is differentiable (see also Lemma 15.14 on 405).

(a) Show for two functions $f, g \in C^1(\mathbb{T}^2)$ and with $\tilde{T} := (\hat{T}^\top)^{-1}$ that

$$\langle f, U^n g\rangle = \langle f, \mathbb{1}\rangle\, \langle \mathbb{1}, g\rangle \;+ \sum_{k\in\mathbb{Z}^2\setminus\{(0,0)\}} f_k\, \overline{g_{\tilde{T}^n k}}\,.$$

(b) Obtain the existence of an $((f,g)$-dependent) constant $L \in (0,\infty)$ for which

$$\left| \langle f, U^n g\rangle - \langle f, \mathbb{1}\rangle \langle \mathbb{1}, g\rangle \right| \le L \left(\sum_{k\in\mathbb{Z}^2\setminus\{0\}} \left(\|k\|_2 + \|\tilde{T}^n k\|_2\right)^{-4}\right)^{\frac{1}{4}}.$$

Hint: Use (9.3.7), the Hölder inequality and $xy \ge \frac{x+y}{2}$ for $x \ge y \ge 1$.

(c) The matrix $\tilde{T}$ is hyperbolic and decomposes the space $\mathbb{R}^2$ into two eigenspaces: $\mathbb{R}^2 = E^s \oplus E^u$. We introduce on $\mathbb{R}^2$ the norm

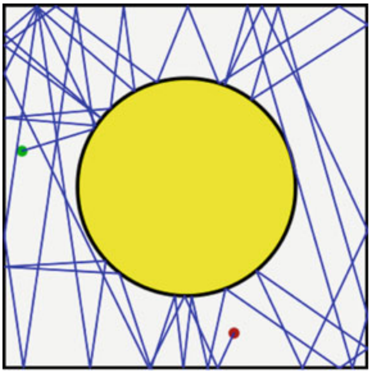
$$\|k\|_E := \|k_s\|_2 + \|k_u\|_2 .$$

Of course, $\|\cdot\|_E$ is equivalent[13] to $\|\cdot\|_2$. But $\|\cdot\|_E$ is easier for calculation, because $\|\tilde{T}k\|_E = \lambda^{-1}\|k_s\|_2 + \lambda\|k_u\|_2$ for $k = k_s + k_u \in E^s \oplus E^u$, with the largest absolute value $\lambda > 1$ of an eigenvalue of $\tilde{T}$.
Show that the correlation decays exponentially[14]:

$$\left|\langle f, U^n g\rangle - \langle f, \mathbb{1}\rangle \langle \mathbb{1}, g\rangle \right| \le C^2 L\Big(\sum_{h\in\mathbb{Z}^2\setminus\{0\}} \|h\|_2^{-4}\Big)^{\frac{1}{4}} \lambda^{-\lfloor \frac{n}{2}\rfloor} \quad (n \in \mathbb{N}). \qquad \diamondsuit$$

9.24 Literature In the example of hyperbolic torus automorphisms, we took advantage of the fact that the torus $\mathbb{R}^2/\mathbb{Z}^2$ is an abelian group, and the mappings are group automorphisms.

In applications in physics, this can of course not be expected, and other techniques will need to be used. An important class of examples for mixing Hamiltonian systems is the geodesic flow on compact manifolds with negative sectional curvature.
The proof that these flows are ergodic can be found in the appendix (written by M. Brin) to the book [Ball] by Ballmann.
Another class of examples consists of *billiards*.
The best known among these is the *Sinai billiard*. In this example, the billiard ball moves on a 2-dimensional torus (or equivalently in a square, see the figure and is reflected at a circular obstacle. This example was first studied by Yakov Sinai in [Sin].



See Cornfeld, Fomin and Sinai [CFS], Kozlov and Treshchev [KT], Liverani and Wojtkowski [LW], and Tabachnikov [Ta]. $\diamondsuit$

[13] i.e., $\exists C \ge 1 : C^{-1}\|k\|_E \le \|k\|_2 \le C\|k\|_E$.

[14] Exponential decay of the correlation is the subject of the book [Bala] by Viviane Baladi.

9.25 Exercise (Product Measure on the Shift Space)
We define subsets called *cylinder sets* of the shift space $M := \mathcal{A}^{\mathbb{Z}}$ over the alphabet $\mathcal{A}$ (see Example 2.18) for $k \in \mathbb{Z}$, $j \in \mathbb{N}_0$, and $\tau = (\tau_1, \ldots, \tau_j) \in \mathcal{A}^j$ by

$$[\tau_1, \ldots, \tau_j]_k^{k+j-1} := \left\{ a : \mathbb{Z} \to \mathcal{A} \;\middle|\; \forall \ell \in \{1, \ldots, j\} \colon a_{k+\ell-1} = \tau_\ell \right\}.$$

So the lower index designates the position of the first determined symbol, the upper index the position of the last one. The cylinder sets generate the σ-algebra (see Definition 9.1.1)

$$\mathcal{M} := \sigma\left(\left\{ [\tau_1, \ldots, \tau_j]_k^{k+j-1} \;\middle|\; k \in \mathbb{Z},\ j \in \mathbb{N}_0,\ \tau_1, \ldots, \tau_j \in \mathcal{A} \right\} \right).$$

For a probability function $p : \mathcal{A} \to [0, 1]$, $\sum_{a \in \mathcal{A}} p(a) = 1$ on the alphabet $\mathcal{A}$, a probability measure μ_p is determined on $\mathcal{M}$ by

$$\mu_p \left([\tau_1, \ldots, \tau_j]_k^{k+j-1} \right) := \prod_{\ell=1}^{j} p(\tau_\ell)$$

according to Kolmogorov's extension theorem (see e.g. KLENKE [Kle]); this measure is called the *product measure*.[15] This measure is invariant under the shift map $\Phi : \mathbb{Z} \times M \to M$.
Show that for all probability functions p, the measure preserving dynamical system $(M, \mathcal{M}, \mu_p, \Phi)$ is mixing. ♢

9.4 Birkhoff's Ergodic Theorem

This theorem by George David Birkhoff studies the existence of the *time average*

$$\overline{f}(m) := \lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} f \circ \Phi_t(m) \qquad (m \in M) \tag{9.4.1}$$

of functions $f : M \to \mathbb{C}$ on the phase space M of a measure preserving dynamical system $(M, \mathcal{M}, \mu, \Phi)$. We will assume here, for simplicity, that μ is a probability measure and Φ is a discrete dynamical system. This dynamical system is thus generated by $\Phi_1 : M \to M$. We introduce the abbreviation $f_t := f \circ \Phi_t$. The Cesàro-means of the sequence of functions $(f_t)_{t \in \mathbb{N}_0}$ will be denoted as[16]

[15] For $\mathcal{A} = \{-1, 1\}$ and $p(1) = t \in [0, 1]$, hence $p(-1) = 1 - t$, this measure is also called the *Bernoulli measure with parameter* t.
[16] The notation A_n for *average* and $S_n f := \sum_{t=0}^{n-1} f_t$ for *sum* are customary in ergodic theory.

$$A_n = A_n f : M \to \mathbb{C} \quad , \quad A_0 := 0 \quad \text{and} \quad A_n := \frac{1}{n}\sum_{t=0}^{n-1} f_t \quad (n \in \mathbb{N}). \tag{9.4.2}$$

So we are asking in (9.4.1) for the existence of the pointwise limit of the sequence of functions $(A_n)_{n\in\mathbb{N}}$. Birkhoff's theorem (Theorem 9.32) answers this question and is therefore also called the *pointwise ergodic theorem*.

(9.4.1) defines the temporal mean $\overline{f}^+(m) := \overline{f}(m)$ of the future. Similarly,

$$\overline{f}^-(m) := \lim_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n-1} f_{-t}(m) \qquad (m \in M)$$

is the temporal mean of the past.

Even if f is continuous, $\overline{f}(m)$ does not exist in general; and if the means $\overline{f}^\pm(m)$ do exist, they need not be equal:

9.26 Examples **(Shift Space)** The sequence space $M := \mathcal{A}^{\mathbb{Z}}$ over the alphabet $\mathcal{A} := \{-1, 1\}$ together with the shift map

$$\Phi_t : M \to M \quad , \quad \big(\Phi_t(a)\big)_k := a_{k+t} \qquad (t \in \mathbb{Z})$$

forms a continuous dynamical system with respect to the product topology on M (see Example 2.18.3). The function

$$f : M \to \mathbb{R} \quad , \quad f(a) := a_0$$

is also continuous with respect to the product topology on M. For instance at location

$$m \in M \quad , \quad m_k := \begin{cases} 1 & , \ k = 0 \\ (-1)^{\lfloor \log_2 |k| \rfloor} & , \ k \in \mathbb{Z}\setminus\{0\} \end{cases} ,$$

where $\lfloor\cdot\rfloor$ denotes the *floor* function, i.e.,

$$m = (\ldots, -1, 1, 1, \overset{k=0}{1}, 1, -1, -1, 1, 1, 1, 1, -1, -1, -1, -1, -1, -1, -1, -1, \ldots),$$

the limit 9.4.1 does not exist, because for $n := 2^l$, the subsequence of Cesàro means (indexed by l)

$$A_n(m) = 2^{-l}\left(1 + \sum_{k=0}^{l-1}(-2)^k\right) = \tfrac{1}{3}\left(2^{2-l} - (-1)^l\right) \qquad (l \in \mathbb{N})$$

$A_n = A_n f$ has the two cluster points $\pm\frac{1}{3}$.

Some deliberation shows that the time average $\overline{f}$ even fails to exist on a subset that is dense in M. Namely, let $\tilde{m} \in M$. Then the sequence of points $x^{(s)} \in M$ with

$$x_k^{(s)} := \begin{cases} \tilde{m}_k \;, \; |k| \le s \\ m_k \;, \; |k| > s \end{cases} \qquad (s \in \mathbb{N},\; k \in \mathbb{Z}) \tag{9.4.3}$$

converges to $\tilde{m}$. However, $\overline{f}\big(x^{(s)}\big)$ does not exist for any of these points.

For example, for the point

$$m \in M \quad , \quad m_k := \begin{cases} 1 & , \; k = 0 \\ \mathrm{sign}(k) & , \; k \in \mathbb{Z} \setminus \{0\} \end{cases} ,$$

one has $\overline{f}^{+}(m) = 1$, but $\overline{f}^{-}(m) = -1$. Here, too, one can find a dense subset of the phase space M, by means of a construction analogous to (9.4.3), such that on this subset, the time averages $\overline{f}^{\pm}(x)$ exist, but are different. ◊

9.27 Exercise (Shift Space) For the function $f : M \to \mathbb{R}$, $m \mapsto m_0$ from Example 9.26, find a dense subset $U \subset M = \{-1, 1\}^{\mathbb{Z}}$, such that for all $m \in U$, the set of cluster points of the sequence $\big(A_n f(m)\big)_{n\in\mathbb{N}}$ is the interval $[-1, 1]$. ◊

The preceding examples may lead to the wrong conjecture that time averages typically do not exist. In any case, if $\overline{f}$ exists, it is invariant under Φ:

9.28 Lemma (Time Averages)
If the time average $\overline{f}(m)$ of $f : M \to \mathbb{C}$ exists for the phase space point $m \in M$, then the same is true for all points of the orbit through m, and

$$\overline{f}\big(\Phi_t(m)\big) = \overline{f}(m) \qquad (t \in \mathbb{Z}).$$

Proof: From the case $t = 1$, the claim for arbitrary $t \in \mathbb{Z}$ follows by induction. Now for $t = 1$, the difference of the Cesàro means for the two initial points equals

$$A_n\big(\Phi_1(m)\big) - A_n(m) = \frac{1}{n}\big(f\,(\Phi_n(m)) - f(m)\big) \qquad (n \in \mathbb{N}). \tag{9.4.4}$$

In the limit $n \to \infty$, the second term $f(m)/n$ on the right converges to zero. While the numerator $f_n(m)$ of the first term $f\big(\Phi_n(m)\big)/n$ will in general be unbounded in n, the assumed convergence of the sequence $\big(A_n(m)\big)_{n\in\mathbb{N}}$ and the equality

$$\frac{f_n(m)}{n} = \left(1 + \frac{1}{n}\right) A_{n+1}(m) - A_n(m)$$

assure nevertheless that the first term on the right side of (9.4.4) converges to zero as well. □

9.29 Remarks (Time Averages)

1. The *existence* of the time average $\overline{f}(m)$ is independent of the choice of a Φ-invariant probability measure μ on M. However, it is by means of μ that we can define what it means for $\overline{f}(m)$ to exist *typically*, namely it means to exist μ-almost everywhere.

2. If for such μ, the dynamical system is ergodic, then, from Birkhoff's ergodic theorem, it follows for integrable functions $f : M \to \mathbb{C}$ with expected value $\mathbb{E}(f) = \int_M f \, \mathrm{d}\mu$ that

$$\overline{f}(m) = \mathbb{E}(f) \qquad \text{for } \mu\text{-almost every } m \in M\,, \tag{9.4.5}$$

in other words the equality of time and space average. If one is ready to believe that the limit $\overline{f}(m)$ *exists* μ-almost everywhere, then formula (9.4.5) is easy to understand.
Namely, for real valued f, if $\overline{f}$ were not constant μ-almost everywhere, then one could obtain a Φ-invariant set of the form $U := \overline{f}^{-1}\big([a,\infty)\big)$ in M, with $\mu(U) \in (0,1)$, contradicting ergodicity. But the constant value cannot be anything but $\mathbb{E}(f)$, because $\mathbb{E}(\overline{f}) = \mathbb{E}(A_n f) = \mathbb{E}(f)$. ◊

As Birkhoff's ergodic theorem does not assume the measure preserving dynamical system to be ergodic, it needs to allow for the following generalization of the right hand side of (9.4.5).

$\mathcal{M}$ denotes σ-algebra of the measurable sets in the phase space M, and in (9.2.1), we introduced the σ-subalgebra $\mathcal{I} \subset \mathcal{M}$ of Φ-invariant measurable sets.

9.30 Definition For the random variable $f \in L^1(M,\mu)$ on the probability space $(M, \mathcal{M}, \mu)$ and the σ-subalgebra $\mathcal{N} \subset \mathcal{M}$, the function $g \in L^1(M,\mu)$ is called a **conditional expectation of** f, **given** $\mathcal{N}$, if the following conditions hold:

1. g is $\mathcal{N}$-measurable (that is $g^{-1}(B) \in \mathcal{N}$ for all Borel sets $B \subseteq \mathbb{C}$),
2. $\mathbb{E}(g 1\!\!1_A) = \mathbb{E}(f 1\!\!1_A)$ for all $A \in \mathcal{N}$.

As a matter of fact, such a conditional expectation of f given $\mathcal{N}$ always exists, and two such conditional expectations differ only on a μ-null set, see for example Klenke [Kle], §8.2. We call them *versions* of the conditional expectations, and write

$$\mathbb{E}(f \mid \mathcal{I})\,.$$

9.31 Example (Conditional Expectation)
For the parabolic torus automorphism given by the matrix $T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ (see Example 9.21 and Figure 9.3.3), the 'horizontal circles' (see Figure 9.4.1)

$$K_r := \left\{ \begin{pmatrix} x \\ r \end{pmatrix} \in \hat{\mathbb{T}}^2 \;\middle|\; x \in [0,1) \right\} \qquad (r \in [0,1))$$

are Φ-invariant, and Φ operates on K_r by rotation with the parameter r. If r is irrational, then every orbit on K_r is dense, whereas for $r \in \mathbb{Q}$, every orbit on K_r is periodic (compare with Example 9.15).

Now let $f : \hat{\mathbb{T}}^2 \to \mathbb{C}$ be a function (say, continuous) with Fourier representation $f = \sum_{k \in \mathbb{Z}^2} c_k \, e_k$ in the orthonormal basis $(e_k)_{k \in \mathbb{Z}^2}$ defined in (9.3.5), with coefficients $c_k \in \mathbb{C}$.

For the σ-subalgebra $\mathcal{I} \subset \mathcal{M}$ of Φ-invariant measurable sets, the function obtained from averaging over the circles K_r,

$$\sum_{k=\binom{0}{k_2}\in\mathbb{Z}^2} c_k e_k : \hat{\mathbb{T}} \to \mathbb{C}, \tag{9.4.6}$$

is a version of the conditional expectation of f given $\mathcal{I}$ (see Figure 9.4.1).

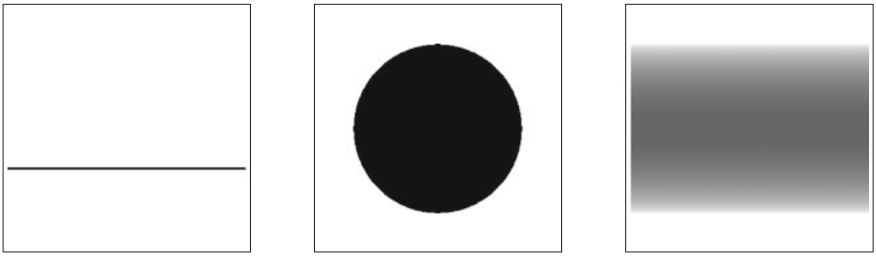


Figure 9.4.1 Invariant circle $K_{1/3}$ (left). Gray scale images of the characteristic function $1\!\!1_D : \hat{\mathbb{T}}^2 \to \mathbb{R}$ of a disc D (center) and its conditional expectation $\mathbb{E}(1\!\!1_D \mid \mathcal{I})$, given the algebra of invariance $\mathcal{I}$ for the torus automorphism generated by $\binom{1\,1}{0\,1}$ (right).

Another version of $\mathbb{E}(f \mid \mathcal{I})$ (which equals the time average $\overline{f}$ of f in *all* points) is given by averaging f over the circles K_r with irrational r, but averages only over the finite period q for circles whose parameter $r = p/q$ is rational. In contrast to the version (9.4.6), this version is in general discontinuous. ◊

s

9.32 Theorem (Birkhoff's Ergodic Theorem, [Bi1])
If $f \in L^1(M,\mu)$, then the time average $\overline{f}(m)$ from (9.4.1) exists for μ-almost every initial value $m \in M$, and $\overline{f} \in L^1(M,\mu)$ and

$$\overline{f} = \mathbb{E}(f \mid \mathcal{I}) \qquad (\mu\text{-almost everywhere}). \tag{9.4.7}$$

9.33 Remark (Birkhoff's Ergodic Theorem)

1. From Lemma 9.28 and $\mathbb{E}(A_n) = \mathbb{E}(f)$, it follows

$$\int_M \overline{f}\, d\mu = \int_M f\, d\mu \quad \text{and} \quad \overline{f} \circ \Phi_t = \overline{f} \qquad (\mu\text{-almost everywhere}).$$

   Moreover, a statement analogous to (9.4.7) holds for $\overline{f}^{\,-}$, so that future and past look similar:

$$\overline{f}^{\,+} = \overline{f}^{\,-} \qquad (\mu\text{-almost everywhere}).$$

2. By considering real and imaginary parts separately, we may assume without loss of generality that f is real valued, $f \in L^1_{\mathbb{R}}(M,\mu)$. Then (with the Cesàro means

$A_n f$ from (9.4.2)), the convergence claim of the ergodic theorem reads:

$$\limsup_{n\to\infty} A_n f = \liminf_{n\to\infty} A_n f \qquad (\mu\text{-almost everywhere}).$$

The advantage of writing it in this form is that lim sup and lim inf are always guaranteed to exist, unlike the limit. The lim sup will be controlled by the maximum, using the following lemma (see for instance [Kle], Chapter 20.2, and KRENGEL [Kre], Chapter 1.2 for a generalization to positive contractions). ◊

9.34 Lemma (Maximal Ergodic Lemma)
For the function $g \in L^1_{\mathbb{R}}(M, \mu)$ on the phase space M of the measure preserving dynamical system $(M, \mathcal{M}, \mu, \Phi)$ and $S_n := \sum_{t=0}^{n-1} g_t$, $S_0 := 0$, let

$$F_n := \left\{m \in M \mid \max\bigl(S_1(m), \ldots, S_n(m)\bigr) > 0\right\}.$$

Then

$$\mathbb{E}(g\, \mathbb{1}_{F_n}) \geq 0 \qquad (n \in \mathbb{N}).$$

Proof: For all $k \in \mathbb{N}_0$, one has $g = S_{k+1} - S_k \circ T$. With $M_n := \max(S_0, \ldots, S_n)$, it follows for $T := \Phi_1$ that

$$g \geq S_{k+1} - M_n \circ T \qquad (k = 0, \ldots, n),$$

hence

$$\begin{aligned}\mathbb{E}(g\, \mathbb{1}_{F_n}) &\geq \mathbb{E}\bigl(\bigl(\max(S_1, \ldots, S_n) - M_n \circ T\bigr)\, \mathbb{1}_{F_n}\bigr) = \mathbb{E}\bigl((M_n - M_n \circ T)\, \mathbb{1}_{F_n}\bigr)\\ &\geq \mathbb{E}\bigl(M_n - M_n \circ T\bigr) = \mathbb{E}\bigl(M_n\bigr) - \mathbb{E}\bigl(M_n\bigr) = 0\,.\end{aligned}$$

Here the first equation follows from $F_n = \{m \in M \mid M_n(m) > 0\}$, the second last from the T-invariance of the measure μ. The second *in*equality follows from $M_n \circ T \geq 0$ and $M_n(m) = 0$ for $m \in M \setminus F_n$ (because $S_0 = 0$). □

Proof of Theorem 9.32: (See Remark 9.33.2 for the proof strategy)

- It suffices to show for all $\varepsilon > 0$ and

$$\begin{aligned}F_\varepsilon^+ := F_\varepsilon^+(f) &:= \left\{m \in M \;\middle|\; \limsup_{n\to\infty} A_n f(m) > \mathbb{E}(f \mid \mathcal{I})(m) + \varepsilon\right\},\\ F_\varepsilon^- := F_\varepsilon^-(f) &:= \left\{m \in M \;\middle|\; \liminf_{n\to\infty} A_n f(m) < \mathbb{E}(f \mid \mathcal{I})(m) - \varepsilon\right\}\end{aligned}$$

  that $\mu(F_\varepsilon^\pm) = 0$. This however already follows from $\mu(F_\varepsilon^+) = 0$ alone, because

$$F_\varepsilon^-(f) = F_\varepsilon^+(-f)\,.$$

- For $\tilde{f} := f - \mathbb{E}(f \mid \mathcal{I}) - \varepsilon$, one has $\mathbb{E}(\tilde{f} \mid \mathcal{I}) = -\varepsilon$, because conditional expectation is idempotent: $\mathbb{E}\big(\mathbb{E}(f \mid \mathcal{I}) \,\big|\, \mathcal{I}\big) = \mathbb{E}(f \mid \mathcal{I})$. Therefore

$$F_\varepsilon^+(f) = F_\varepsilon^+(\tilde{f}) = \left\{ m \in M \;\middle|\; \limsup_{n\to\infty} A_n \tilde{f}(m) > 0 \right\}. \tag{9.4.8}$$

- We now let $g := \tilde{f}\, \mathbb{1}_{F_\varepsilon^+}$. By the Φ-invariance of $\limsup_{n\to\infty} A_n g$, it follows that $F_\varepsilon^+ \in \mathcal{I}$, and therefore the iterates $g_k = g \circ \Phi_k$ are of the form $g_k = \tilde{f}_k\, \mathbb{1}_{F_\varepsilon^+}$.
  As in the proof of Lemma 9.34, we use

$$M_n\, g := \max(S_0 g, \ldots, S_n g) \quad \text{and} \quad F_n := \{m \in M \mid M_n g(m) > 0\}.$$

  Since $n \mapsto M_n$ is increasing, $F_n \subseteq F_{n+1}$, and with (9.4.8),

$$\begin{aligned}
\bigcup_{n\in\mathbb{N}} F_n &= \{m \in M \mid \lim_{n\to\infty} M_n g(m) > 0\} \\
&= \left\{ m \in M \;\middle|\; \sup_{n\in\mathbb{N}} S_n g(m) > 0 \right\} = \left\{ m \in M \;\middle|\; \sup_{n\in\mathbb{N}} A_n g(m) > 0 \right\} \\
&= \left\{ m \in M \;\middle|\; \sup_{n\to\infty} A_n \tilde{f}(m) > 0 \text{ and } \limsup_{n\to\infty} A_n \tilde{f}(m) > 0 \right\} = F_\varepsilon^+(\tilde{f}).
\end{aligned}$$

- Pointwise convergence of the functions $g\, \mathbb{1}_{F_n}$ to $g\, \mathbb{1}_{F_\varepsilon^+} = g$ and their domination by $|g|$ follows by the dominated convergence theorem

$$\lim_{n\to\infty} \mathbb{E}(g\, \mathbb{1}_{F_n}) = \mathbb{E}(g).$$

  By the maximal ergodic lemma 9.34, $\mathbb{E}(g\, \mathbb{1}_{F_n}) \geq 0$, hence also $\mathbb{E}(g) \geq 0$.
- On the other hand, with $F_\varepsilon^+ \in \mathcal{I}$ and $\mathbb{E}(\tilde{f} \mid \mathcal{I}) = -\varepsilon$, one obtains

$$\mathbb{E}(g) = \mathbb{E}(\tilde{f}\, \mathbb{1}_{F_\varepsilon^+}) = \mathbb{E}\left(\mathbb{E}(\tilde{f} \mid \mathcal{I})\, \mathbb{1}_{F_\varepsilon^+}\right) = -\varepsilon\, \mathbb{E}(\mathbb{1}_{F_\varepsilon^+}) = -\varepsilon\, \mu(F_\varepsilon^+),$$

  which is compatible with $\mathbb{E}(g) \geq 0$ only when $\mu(F_\varepsilon^+) = 0$. □

9.35 Exercise (Birkhoff's Ergodic Theorem for Flows)
Let $(M, \mathcal{M}, \mu, \Phi)$ be a measure preserving dynamical system with a compact metric space M, the Borel σ-algebra $\mathcal{M}$, a probability measure μ, and continuous flow $\Phi : \mathbb{R} \times M \to M$. Moreover, let $f \in L^1(M, \mu)$. Show that the time average

$$\overline{f}(m) := \lim_{t\to\infty} \frac{1}{t} \int_0^t f \circ \Phi(s, m)\, \mathrm{d}s$$

exists for μ-almost all $m \in M$, and that moreover $\overline{f} \in L^1(M, \mu)$, $\int_M \overline{f}\, \mathrm{d}\mu = \int_M f\, \mathrm{d}\mu$, and $\overline{f}(m) = \overline{f} \circ \Phi(t, m)$ for μ-almost all $m \in M$ and all $t \in \mathbb{R}$.

Hint: Show first that $\int_0^n f \circ \Phi(s, m)\, \mathrm{d}s = \sum_{k=0}^{n-1} F \circ T^k(m)$ for $n \in \mathbb{N}, T := \Phi(1, \cdot)$, and $F(m) := \int_0^1 f \circ \Phi(s, m)\, \mathrm{d}s$. $\diamondsuit$

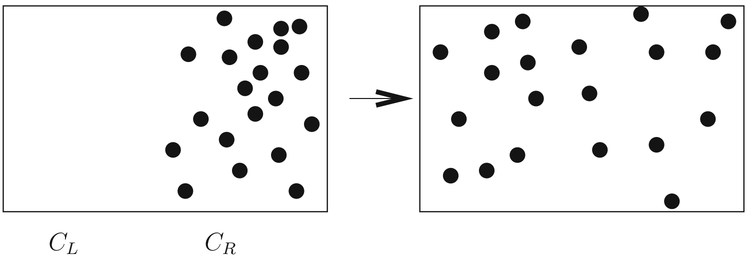
9.36 Exercise ((Normal Real Numbers)
Show that for Lebesgue-almost all $x \in [0, 1)$, the frequency with which the digit 1 occurs in the binary representation of x is $\frac{1}{2}$.
Hint: Consider the mapping $T(x) := 2x \pmod 1$ on $[0, 1)$ and use Birkhoff's ergodic theorem 9.32 with an appropriate observable. $\diamondsuit$

9.37 Remark (Measure Theoretically Typical Dynamics)
For a given dynamical system $\Phi : \mathbb{Z} \times M \to M$, there are in general many Φ-invariant probability measures on the phase space M.

For example, if $m \in M$ is a periodic point with period T, the probability measure $\frac{1}{T} \sum_{t=0}^{T-1} \delta_{\Phi_t(m)}$ is invariant and ergodic, and so are the product measures from Exercise 9.25. Accordingly, the typical behavior predicted by the ergodic theorem will strongly depend on the choice of the measure. $\diamondsuit$

9.5 Poincaré's Recurrence Theorem

Imagine a container $C = C_L \dot{\cup} C_R \subset \mathbb{R}^3$ divided into two chambers by a wall. Assume the left side C_L to hold a vacuum, whereas the right side C_R is filled with air. We now remove the dividing wall. Without knowing the precise positions and velocities of the gas molecules, we expect the molecules not to stay in the right half, but to spread out about evenly over both sides.



Nobody would expect, intuitively, that at some later time, the air would regather in the right half of the container. However, this is what is going to happen as a consequence of the following theorem:

9.38 Theorem (Poincaré Recurrence Theorem, [Poi2]**)**
Let $\Phi : \mathbb{Z} \times M \to M$ be a dynamical system that preserves the measure μ on M (see Definition 9.5). Moreover, assume that $B \subseteq \widetilde{M} \subseteq M$ are measurable and $\widetilde{M}$ is Φ-invariant, with $\mu(\widetilde{M}) < \infty$.

Then μ-almost all points of B return to B infinitely often.

9.39 Remarks (Poincaré Recurrence Theorem)

1. If $\mu(M) < \infty$, one can simply set $\widetilde{M} := M$. Without the hypothesis $\mu(\widetilde{M}) < \infty$, however, the theorem does not hold, as the simple example of a translation $\Phi_t(x) := x + t$ on $M := \mathbb{R}$ with the Lebesgue measure μ shows.
2. If we restrict a flow $\Phi : \mathbb{R} \times M \to M$ to the discrete times $t \in \mathbb{Z}$, we can apply the theorem to the restricted system.
   In the case of discrete time, it is also more natural to formalize the colloquial statement "x returns to B infinitely often" for $x \in B$ as

$$\big|\{n \in \mathbb{N} \mid \Phi_n(x) \in B\}\big| = \infty . \qquad \diamondsuit$$

Proof of Theorem 9.38:

- If $\mu(B) = 0$, the claim is obvious, since in measure theory, 'almost all' means 'all with the exception of a set of measure zero'.

- So assume $\mu(B) > 0$ and let $K_n := \bigcup_{j=n}^{\infty} \Phi_{-j}(B) \quad (n \in \mathbb{N}_0)$.
  K_n is a countable union of measurable sets, hence measurable, and $\widetilde{M} \supseteq K_n$ implies $\mu(K_n) \le \mu(\widetilde{M}) < \infty$. The sets K_n satisfy

$$K_{n+1} = \Phi_{-1}(K_n) \quad \text{and} \quad \widetilde{M} \supset K_0 \supseteq K_1 \supseteq \ldots \supseteq K_n \supseteq K_{n+1} \supseteq \ldots .$$

  $B \cap K_n$ is the set of those points from B that return to B after time n.

- $B \cap (\bigcap_{n\in\mathbb{N}_0} K_n)$ is the set of those points from B that keep returning to B after arbitrarily long times, i.e., that return infinitely often. This set is measurable, because it is the countable intersection of measurable sets. We now want to show that

$$\mu\big(B \cap (\textstyle\bigcap_{n\in\mathbb{N}_0} K_n)\big) = \mu(B) . \tag{9.5.1}$$

  Since the K_n are nested,

$$\bigcap_{n\in\mathbb{N}_0} K_n = K_0 \Big\backslash \Big(\dot{\bigcup}_{n\in\mathbb{N}_0} K_n \setminus K_{n+1}\Big) ,$$

  and therefore by σ-additivity (9.1.2) of μ, one gets

$$\mu\big(B \cap \big(\bigcap_{n\in\mathbb{N}_0} K_n\big)\big) = \mu(B \cap K_0) - \sum_{n=0}^{\infty} \mu\big(B \cap (K_n \setminus K_{n+1})\big) . \tag{9.5.2}$$

  But $B \cap K_0 = B$; and from $K_n \supseteq K_{n+1}$ and $\mu(K_{n+1}) = \mu\big(\Phi_{-1}(K_n)\big) = \mu(K_n)$, it follows that $\mu(K_n \setminus K_{n+1}) = 0$. Therefore, (9.5.2) implies formula (9.5.1).

$\square$

9.40 Examples (Statistical Mechanics and Reversibility)
In the example of the container $C \subset \mathbb{R}^3$, $(\mathbb{R}^3 \times C)^n$ is the phase space of the n atoms of some gas. We model the interaction between the atoms with a smooth function $V \in C^\infty(\mathbb{R}^3, \mathbb{R})$. Then the Hamiltonian is of the form

$$H : (\mathbb{R}^3 \times C)^n \to \mathbb{R} \quad , \quad H(p_1, q_1, \ldots, p_n, q_n) = \sum_{k=1}^{n} \tfrac{1}{2}\|p_k\|^2 + \sum_{1 \le k < l \le n} V(q_l - q_k).$$

We stipulate that in case of hits against the walls of the container, the particles will be reflected (outgoing angle = −incoming angle). This gives us a dynamical system that preserves the Lebesgue measure.[17]

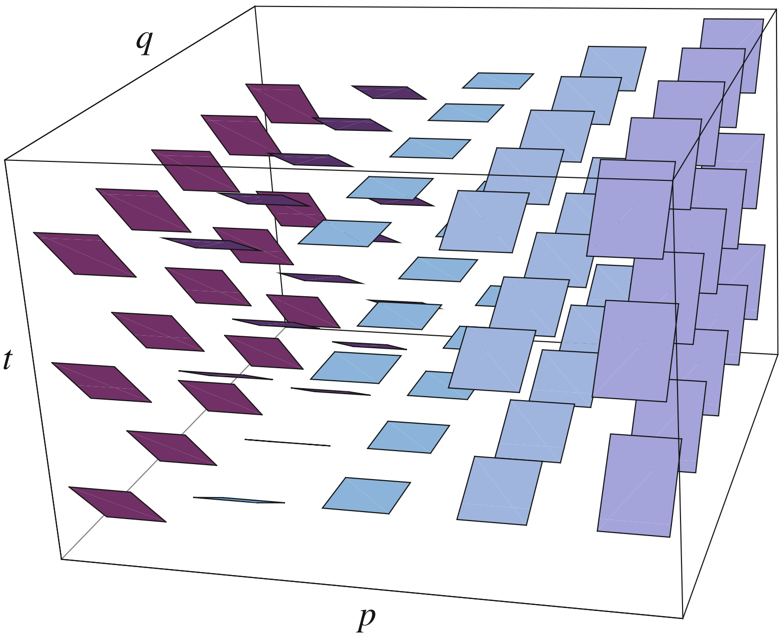
Now consider, for a regular value E of H, the energy shell $\widetilde{M} := M := H^{-1}(E)$ with its (finite) Liouville measure μ, and set $B := M \cap (\mathbb{R}^3 \times C_R)^n$ in Theorem 9.38. This is the portion of M that represents all n particles being located in the right half of the container.

A crucial point with Poincaré's recurrence theorem is that no assertion is made concerning the time of the recurrence. In the case of the container, the recurrence time to B (with realistic parameters for the container C, the number of particles n and energies E) is larger than the present age of the universe. This is essential for the practical validity of statistical mechanics with its claims about irreversibility. ◊

9.41 Literature The theorem just proved belongs to ergodic theory (namely the connection of the theory of dynamical systems with measure and probability theory); likewise Theorem 12.16 belongs here. A good book on ergodic theory is WALTERS [Wa2]; applications to problems of classical mechanics can be found in ARNOL'D and AVEZ [AA], and in BUNIMOVICH [Bu]. ◊

[17] This system is not continuous, but it is a dynamical system in the measure theoretic sense.

Chapter 10
Symplectic Geometry



Contact structure for the contact form $dt - p\,dq$ on the extended phase space $\mathbb{R}^3$

When one leaves the special case of linear Hamiltonian differential equations behind, the symplectic bilinear form studied in Chapter 6 becomes a symplectic form, and Lagrangian subspaces become Lagrangian submanifolds. Symplectic manifolds, i.e., manifolds with a symplectic form, have a special kind of geometry. Their structure preserving mappings are called canonical.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_10

10.1 Symplectic Manifolds

The dynamical systems studied so far frequently had as their phase spaces open subsets of $\mathbb{R}^n$. Starting with this chapter, we will systematically study more general phase spaces that are differentiable manifolds: (see Appendices A.2 and A.3). There are three reasons to embrace this generality:

- Firstly, the existence of *constants of motion* in mechanics leads to (sub-) manifolds in a natural way.
- Secondly, the *geometric contents* of statements is displayed more clearly if they are made (as far as possible) for manifolds and not just $\mathbb{R}^n$. This is because in the latter case, one usually ends up calculating in arithmetic coordinates; and moreover, $\mathbb{R}^n$ has additional structures (it is a vector space, has a canonical metric, etc.), from which we should rather abstract.
- Furthermore, this generalization requires little effort, which to some extent has been made already (for instance in Lemma 8.18 for the Euler-Lagrange equation on tangent bundles of manifolds).

The Hamiltonian formalism uses as phase space the cotangent bundle T^*M of a manifold M, rather than its tangent bundle TM.
The relation between the two is just like the relation between a vector space and its dual space:

10.1 Definition *Let M be a differentiable manifold.*

- *A 1-form on the tangent space T_xM of M at x (i.e., a linear mapping $T_xM \to \mathbb{R}$) is also called a* **cotangent vector of M at x** .
- *The dual space to T_xM, i.e., the linear space T_x^*M of such 1-forms, is also called the* **cotangent space to M at x**.
- *The union $T^*M := \bigcup_{x\in M} T_x^*M$ is called* **cotangent bundle of M**.

Coordinates. Let $n := \dim(M)$. If local coordinates $q = (q_1, \ldots, q_n) : U \to \mathbb{R}^n$ are given on a neighborhood $U \subseteq M$ of x, the 1-form p on T_xM is determined by the n numbers $p_\ell := p\left(\frac{\partial}{\partial q_\ell}(x)\right)$. With respect to these coordinates, we then identify p with the vector $(p_1, \ldots, p_n)$ and use the bundle coordinates (p, q) on T^*U.

Obviously $\dim(T^*M) = 2\dim(M)$. The cotangent bundle T^*M is even diffeomorphic to TM (but not canonically diffeomorphic; see page 503).

10.2 Remark (Cotangent Bundles as Phase Spaces) We have already encountered some examples of cotangent bundles T^*M, namely

1. $T^*\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$ (free motion)
2. $T^*S^1 \cong \mathbb{R} \times S^1$ (planar pendulum)
3. $T^*\mathbb{T}^2 \cong \mathbb{R}^2 \times \mathbb{T}^2$ (double pendulum).

The Hamilton functions in these examples were smooth mappings $H : T^*M \to \mathbb{R}$. The right side of the Hamiltonian differential equation was given by the Hamiltonian vector field X_H induced by H. If we denote the phase space as $P := T^*M$, then X_H is in particular a smooth mapping $X_H : P \to TP$ into the tangent bundle of P, subject to the property that $\pi_P \circ X_H : P \to P$ is the identity mapping; in other words, X_H is a tangent vector field.

With respect to the local (p, q) bundle coordinates, X_H has the components

$$X_H = \left(-\frac{\partial H}{\partial q_1}, \ldots, -\frac{\partial H}{\partial q_n}\,;\ \frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_n}\right)^\top . \qquad \Diamond$$

The following, coordinate free, definition of the Hamiltonian vector field permits a better geometric insight.

10.3 Definition

- *A* **symplectic form** *on a manifold P is a closed $(d\omega = 0)$ 2-form ω on P that is* **nondegenerate**, *i.e., for all $x \in P$* and *$\xi \in T_x P \setminus \{0\}$,*

$$\exists \eta \in T_x P : \quad \omega(\xi, \eta) \neq 0 .$$

- *(P, ω) is then called a* **symplectic manifold**.

10.4 Remark

1. It follows from linear algebra (see Theorem 6.13.2) that $\dim(T_x P)$ must be even for $x \in P$ in order for the 2-form to be nondegenerate. Of course, phase spaces P of the form $P = T^*M$ meet this condition.
2. Not every closed k-form α (i.e., $d\alpha = 0$) is *exact* (i.e., $\alpha = d\beta$ for some $(k-1)$-form β). If ω is exact, we call (P, ω) *exact symplectic*.
3. In the case $P = T^*\mathbb{R}^n$, the 2-form $\omega_0 := \sum_{i=1}^n dq_i \wedge dp_i$ is available. $(T^*\mathbb{R}^n, \omega_0)$ is exact symplectic, since $\omega_0 = -d\theta_0$ with $\theta_0 := \sum_{i=1}^n p_i dq_i$. $\Diamond$

We know that an arbitrary (not necessarily antisymmetric) nondegenerate bilinear form allows to transition from vector fields to 1-forms and back. If we apply this to the symplectic 2-form, the relation is given by the equation

$$\omega(X, \cdot) = \alpha \qquad X\ , \ \text{ a vector field}\ ,\ \alpha \text{ a 1-form on } P\ .$$

X can be determined from α, and α can be determined from X.

10.5 Definition *A vector field $X : P \to TP$ on the symplectic manifold (P, ω) is called*

- **Hamiltonian** *if $\omega(X, \cdot)$ is an exact 1-form,*
- **locally Hamiltonian** *if the 1-form $\omega(X, \cdot)$ is closed.*

- *For a continuously differentiable function* $H : P \to \mathbb{R}$, *the vector field* X_H *defined by* $\omega(X_H, \cdot) = dH$ *is called the* **Hamiltonian vector field generated by** H, *and* (P, ω, H) *is called a* **Hamiltonian system**.

10.6 Example Considering the case $(P, \omega) = (T^*\mathbb{R}^n, \omega_0)$, one obtains that, with respect to (p, q) coordinates,

$$X_H = \sum_{i=1}^n \left((X_H)_i \frac{\partial}{\partial p_i} + (X_H)_{n+i} \frac{\partial}{\partial q_i} \right)$$

$$\sum_{i=1}^n dq_i \wedge dp_i(X_H, \cdot) = \sum_{i=1}^n \bigl((X_H)_{i+n} dp_i - (X_H)_i dq_i \bigr)$$

and

$$dH = \sum_{i=1}^n \left(\frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i \right),$$

hence by comparing coefficients:

$$(X_H)_i = -\frac{\partial H}{\partial q_i} \quad , \quad (X_H)_{i+n} = \frac{\partial H}{\partial p_i} \qquad (i = 1, \ldots, n), \tag{10.1.1}$$

namely indeed the right hand side of the Hamilton equations. ◊

One may ask why we used specifically the form ω_0. As a matter of fact, in Exercise 10.8, we will encounter a formulation of the motion of a particle in a magnetic field B that uses a different symplectic form $\omega_B \neq \omega_0$.

Nevertheless, ω_0 does play a special role: namely it can be defined geometrically without reference to coordinates. This definition will carry over to all phase spaces P that are cotangent bundles ($P = T^*M$). We denote by

$$\pi_M^* : P \to M \quad , \quad \pi_M^*(T_q^*M) = \{q\} \tag{10.1.2}$$

the *base point projection* of the cotangent bundle $P := T^*M$ of M.

We want to define a 1-form on the phase space P whose exterior derivative gives the canonical symplectic form. Such a 1-form θ_0 is defined by how it maps an arbitrary tangent vector field $Y : P \to TP$. Then $\theta_0(Y) : P \to \mathbb{R}$ is a function on the phase space. Considering a point $x \in P$ in phase space, it can be viewed as a cotangent vector of the configuration manifold M at the base point $q := \pi_M^*(x)$. Such a cotangent vector can in turn be applied to a tangent vector to M at the same point. Combining these observations, we can define a 1-form θ_0 on P by

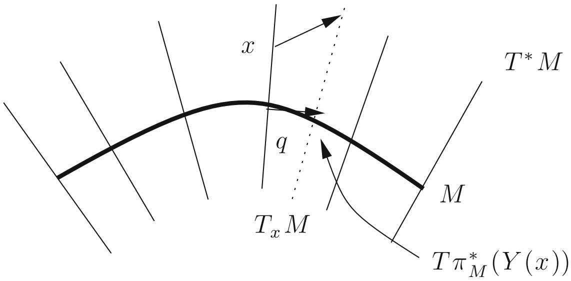


Figure 10.1.1 On the definition of the canonical symplectic form θ_0

$$\langle\theta_0(x), Y(x)\rangle := \big\langle x, T\pi^*_M\big(Y(x)\big)\big\rangle \qquad (x \in P) \tag{10.1.3}$$

(Figure 10.1.1), because the tangent mapping $T\pi^*_M$ maps tangent vectors of phase space P into such of the configuration space M, and these in turn can be paired with the 1-form x to result in a number. The following diagram of manifolds and mappings *commutes*,

$$\begin{array}{ccc} TP & \xrightarrow[T\pi^*_M]{} & TM \\ \pi_P\big\downarrow & & \big\downarrow \pi_M \\ P & \xrightarrow{\pi^*_M} & M \end{array} .$$

i.e., the two concatenated mappings from TP to M are equal to each other.

10.7 Definition

- *The differential form* θ_0 *defined by* (10.1.3) on the *phase space* $P = T^*M$ *is called the* **canonical 1-form**, **Liouville form** *or* **tautological form**.
- $\omega_0 := -d\theta_0$ *is called the* **canonical symplectic form**.

In a canonical bundle chart $(p, q) : T^*U \to \mathbb{R}^n_p \times U$ of T^*M, one has

$$Y = \sum_{i=1}^{n}\left(Y_i\frac{\partial}{\partial p_i} + Y_{i+n}\frac{\partial}{\partial q_i}\right) \quad \text{and} \quad (T\pi^*_M)Y = \sum_{i=1}^{n} Y_{i+n}\frac{\partial}{\partial q_i} .$$

The vector p can be written in the form $p = \sum_{i=1}^n p_i\, dq_i$, so we obtain

$$\theta_0 = \textstyle\sum_{i=1}^n p_i\, dq_i .$$

However, p is a 1-form on M, whereas θ_0 is a 1-form on $P = T^*M$!

So it suffices to specify a (Hamilton) function $H : T^*M \to \mathbb{R}$ on the cotangent bundle of a configuration manifold M to define a Hamiltonian vector field X_H on this phase space by virtue of the relation

$$\omega_0(X_H, \cdot) = dH \, ,$$

and thus to define a differential equation there.

10.8 Exercise (Particles in a Magnetic Field)

Let $B = (B_1, B_2, B_3)^\top \in C^\infty\big(\mathbb{R}^3_q, \mathbb{R}^3\big)$ a location dependent *magnetic field*, i.e., a divergence free vector field, $\mathrm{div}(B) = 0$. The 2-form $\omega_B \in \Omega^2(P)$ on the phase space $P := \mathbb{R}^3_q \times \mathbb{R}^3_v$ will be defined by

$$\omega_B := \omega_0 + B_1\, dq_2 \wedge dq_3 + B_2\, dq_3 \wedge dq_1 + B_3\, dq_1 \wedge dq_2 \text{ with } \omega_0 = \sum_{i=1}^{3} dq_i \wedge dv_i .$$

(a) Show that ω_B is a symplectic form on P.
(b) For vector fields $X, Y : P \to \mathbb{R}^6$, calculate the function $\omega_B(X, Y) : P \to \mathbb{R}$.
(c) For $H : P \to \mathbb{R}$, $H(q, v) := \frac{1}{2}\|v\|^2$, calculate the 1-form dH and $dH(Y)$. Find the unique solution X_H of the equation $\omega_B(X_H, Y) = dH(Y)$, where $Y : P \to \mathbb{R}^6$ are arbitrary vector fields.
(d) Write out the Hamiltonian differential equation $\dot{x} = X_H(x)$ in position and velocity coordinates $x = (q, v) \in P$. ◊

This example shows that one can use symplectic forms on cotangent bundles that differ from the canonical symplectic form ω_0. Moreover, not every symplectic manifold (P, ω) is a cotangent bundle, and not every symplectic manifold is exact symplectic:

10.9 Example (2-Torus) In local (!) angle coordinates φ_1, φ_2, the area form $\omega := d\varphi_1 \wedge d\varphi_2$ is a symplectic form on the 2-torus $P := \mathbb{T}^2$. But $\int_P \omega = 4\pi^2 \neq 0$. So the symplectic manifold (P, ω) is not exact symplectic, because if we had $\omega = -d\Theta$, then by Stokes' Theorem (see Theorem B.39), it would follow

$$\int_P \omega = -\int_P d\Theta = -\int_{\partial P} \Theta = 0 \, ,$$

as the compact manifold $P = \mathbb{T}^2$ does not have a boundary ($\partial P = \emptyset$).

More generally, all orientable and closed (i.e., compact without boundary) surfaces have an area form, which is a symplectic form, but this form is not exact symplectic, for the same reason. ◊

Also, not every locally Hamiltonian vector field is Hamiltonian.

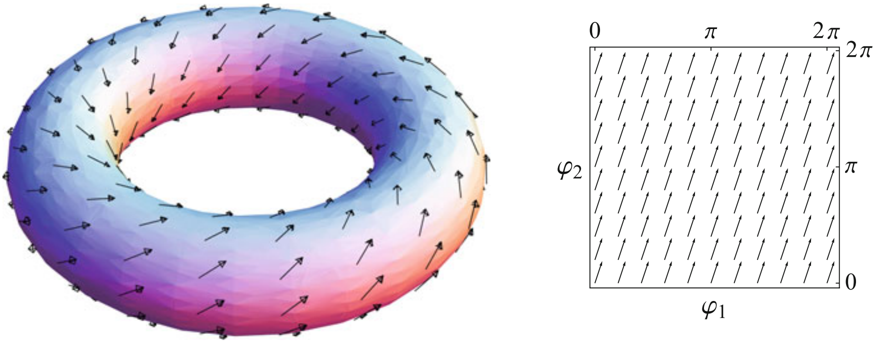


Figure 10.1.2 Locally, but not globally, Hamiltonian vector field on the torus $\mathbb{T}^2$. Left graphic: $\mathbb{T}^2 \subset \mathbb{R}^3$, right graphic: $\mathbb{T}^2 = \mathbb{R}^2/(2\pi\mathbb{Z})^2$

10.10 Example **(Conditionally Periodic Motion on $\mathbb{T}^2$)**
On the phase space $P = \mathbb{T}^2$, define the vector field $X : P \to TP$ in local angle coordinates $\varphi = (\varphi_1, \varphi_2)$ by $\varphi \mapsto (\varphi; a)$, where $a = (a_1, a_2) \in \mathbb{R}^2$.
X is locally Hamiltonian, since the 1-form $\beta := \omega(X, \cdot)$ is closed:

$$\beta = d\varphi_1 \wedge d\varphi_2 \left(a_1 \frac{\partial}{\partial \varphi_1} + a_2 \frac{\partial}{\partial \varphi_2} \right) = a_1\, d\varphi_2 - a_2\, d\varphi_1 \quad \text{, hence} \quad d\beta = 0\,.$$

However, for $a \neq 0$, there does not exist a function $H : \mathbb{T}^2 \to \mathbb{R}$ such that $\beta = dH$, because the 1-form β does not vanish anywhere, whereas a function $H \in C^1(\mathbb{T}^2)$ must take on a maximum at some $x \in \mathbb{T}^2$ on the (compact!) torus, and $dH(x) = 0$ at such a maximum (Figure 10.1.2).

The time-1 flow of X, namely the translation $\varphi \mapsto \varphi + a$, is also not generated by *time dependent* Hamiltonian vector fields, see Example 10.29.3. ◊

10.11 Remark **(Contact Manifolds)**
Symplectic manifolds have an even dimension. There is however a similarly important geometric structure on certain odd-dimensional manifolds .

By definition, a *contact form* on a manifold M of dimension $2n + 1$ is a 1-form $\theta \in \Omega^1(M)$, for which the $(2n + 1)$-form $\theta \wedge (d\theta)^{\wedge n}$ does not vanish anywhere. In this case, (M, θ) is called a *contact manifold.*

1. An example is the extended phase space $M := T^*N \times \mathbb{R}_t$ of a cotangent bundle (T^*N, θ_0), with the Liouville form θ_0 from (10.1.3). If we denote the projections by $\pi_1 : M \to T^*N$ and $\pi_2 : M \to \mathbb{R}_t$, then $\theta := \pi_2^* dt - \pi_1^*\theta_0$ is a contact form. In the case of $N = \mathbb{R}^n_q$, one has therefore $\theta = dt - \sum_{k=1}^n p_k\, dq_k$.
2. If g is a Riemannian metric on N and $H : T^*N \to \mathbb{R}$ is of the form $H(p, q) = \frac{1}{2} g_q(p, p) + V(q)$, and we consider a level set $M := H^{-1}(h)$ for $h > \sup_{q \in N} V(q)$, then the restriction of θ_0 to M is a contact form. Indeed, letting $n := \dim(N)$, the form $dH \wedge \theta_0 \wedge (d\theta_0)^{\wedge n-1}$ is proportional to $(H - V)\,(d\theta_0)^{\wedge n}$, and therefore not degenerate on $M \subset T^*N$.

3. The same is not true for energies h that are smaller than the supremum of the potential V. For example, the restriction of $\theta_0 = \sum_{k=1}^n p_k\, dq_k$ is not a contact form for the energy surfaces $M = H^{-1}(h)$ of the harmonic oscillator

$$H : \mathbb{R}^{2n} \to \mathbb{R} \quad , \quad H(p,q) = \tfrac{1}{2}(\|p\|^2 + \|q\|^2)\,.$$

However, in this case, one can use $\theta := \frac{1}{2}\sum_{k=1}^n (p_k\, dq_k - q_k\, dp_k)$, which is a 1-form that differs from the Liouville form $\theta_0 = \sum_{k=1}^n p_k\, dq_k$ by an exact form: $\theta_0 = \theta + d\langle p,q\rangle/2$. One can prove, analogous to Case 2 above, that θ is a contact form on M.

The kernel of θ defines hyperplanes in the tangential spaces $T_x M$, and the collection of these hyperplanes is a geometric distribution on M, called the *contact distribution*. What is special about θ is that it is maximally non-integrable in the sense of Frobenius (see Appendix F.3). This implies in particular that the contact distribution cannot be tangential to any $2n$-dimensional submanifold of M. This can be seen intuitively in the example of the contact distribution on $\mathbb{R}^3$ depicted on page 215. ◊

10.12 Literature An introduction to contact geometry is given by H.Geiges in [Ge] and by V.I.Arnol'd and A.Givental in [AG]. ◊

10.2 Lie Derivative and Poisson Bracket

"La vie n'est bonne qu'à deux choses: à faire des mathématiques et à les professer." Siméon Poisson[1]

In the previous chapter, we introduced the notion of a symplectic manifold (P,ω). Here, P was a manifold, and ω a symplectic 2-form, i.e., a nondegenerate closed differential form of 2nd degree. The standard example is

$$(P,\omega) = \left(T^*\mathbb{R}^n_q\,,\; \textstyle\sum_{i=1}^n dq_i \wedge dp_i\right).$$

A Hamilton function $H : P \to \mathbb{R}$ will then induce the Hamiltonian vector field X_H, which is defined by the relation $\omega(X_H,\cdot) = dH$. In the standard example, in (p,q) coordinates, one has

$$X_H = \left(-\tfrac{\partial H}{\partial q_1}, \cdots, -\tfrac{\partial H}{\partial q_n}\;;\; \tfrac{\partial H}{\partial p_1}, \cdots, \tfrac{\partial H}{\partial p_n}\right)^{\top}.$$

Any vector field, in particular X_H, generates a flow Φ_t that exists at least for small times.

[1] Translation: "Life is only good for two things: to do mathematics and to teach it".
Quoted after: François Arago: Notices biographiques, Volume 2, 662 (1854).

In mechanics, the relation $\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi_t^*\omega\right) = \Phi_t^* L_X\omega$ between flow and Lie derivative, which was proved in Theorem B.34, has the following consequence:

10.13 Theorem *Let (M, ω, H) be a Hamiltonian system, and let the Hamiltonian vector field X_H generate the flow $\Phi : \mathbb{R} \times M \to M$. Then, for all times $t \in \mathbb{R}$,*

$$\Phi_t^*\, \omega^{\wedge k} = \omega^{\wedge k} \qquad \left(k \in \{1, \ldots, \tfrac{1}{2}\dim M\}\right).$$

Proof

- $\Phi_0 = \mathrm{Id}_M$, hence (by Theorem B.15.3) $\Phi_0^*\, \omega^{\wedge k} = \omega^{\wedge k}$.
- $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t^*\omega = \Phi_t^* L_{X_H}\omega = \Phi_t^*(\mathbf{i}_{X_H} d + d\mathbf{i}_{X_H})\omega = \Phi_t^*\big(\mathbf{i}_{X_H} \underbrace{d\omega}_{0} + \underbrace{ddH}_{0}\big) = 0,$ hence by the Leibniz rule also $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t^*\,\omega^{\wedge k} = \frac{\mathrm{d}}{\mathrm{d}t}(\underbrace{\Phi_t^*\omega \wedge \ldots \wedge \Phi_t^*\omega}_{k \text{ times}}) = 0.$ □

10.14 Remark (Symplectic Forms and Hamiltonian Flows)

1. So Hamiltonian flows leave the symplectic 2-form ω invariant. This is an important property of consistency, since the vector field X_H generating the flow was defined by means of ω. As we have seen, it follows from the requirement that a symplectic form be closed, a requirement that may initially have seemed unmotivated.
   In contrast, for exact symplectic 2-forms (i.e., those of the form $\omega = -d\Theta$), the 1-form Θ is in general not invariant.
2. In particular, the volume form $\omega^{\wedge n}$ with $n = \frac{1}{2}\dim(M)$ is flow invariant. So Hamiltonian flows preserve volume.
   Conversely, on $\mathbb{R}^2$, every volume preserving flow is Hamiltonian, because for the generating vector field X, the form $\omega(X, \cdot)$ will then be closed, hence by Poincaré's lemma (see Appendix B.7), exact.

   On the other hand, for $n > 1$, not every volume preserving flow on $\mathbb{R}^{2n}$ will be Hamiltonian. ◊

An important role in mechanics is played by the Poisson bracket:

10.15 Definition *Let (M, ω) be a symplectic manifold and $f, g \in C^\infty(M, \mathbb{R})$. The* **Poisson bracket** *of f and g is the function*

$$\{f, g\} := \omega(X_f, X_g) \in C^\infty(M, \mathbb{R})\,. \tag{10.2.1}$$

10.16 Theorem $\{f, g\} = -L_{X_f} g = +L_{X_g} f.$

Proof $-L_{X_f} g = -\mathbf{i}_{X_f} dg = -\mathbf{i}_{X_f}\mathbf{i}_{X_g}\omega = -\omega(X_g, X_f) = \omega(X_f, X_g).$ □

This relation with the Lie derivative means that the Poisson bracket $g \mapsto \{f, g\}$ with a given function f is a derivation, i.e., it satisfies the Leibniz rule

$$\{f, gh\} = \{f, g\}h + g\{f, h\} \quad \text{and therefore also} \quad \{f, G \circ h\} = G' \circ h \ \cdot \{f, h\}$$

for continuously differentiable $G : \mathbb{R} \to \mathbb{R}$.

10.17 Theorem *Let $f \in C^\infty(M, \mathbb{R})$ generate a Hamiltonian flow Φ on M. Then the following are equivalent for $g \in C^\infty(M, \mathbb{R})$:*

1. $\{f, g\} = 0$
2. *g is constant on the orbits of Φ.*

Proof By Theorems B.34 and 10.16, we have

$$\tfrac{d}{dt} g \circ \Phi_t = \tfrac{d}{dt} \Phi_t^* g = \Phi_t^* L_{X_f} g = -\Phi_t^*\big(\{f, g\}\big)\,. \qquad \square$$

10.18 Remark (Poisson Bracket)
In **canonical coordinates** $(p_1, \ldots, p_n, q_1, \ldots, q_n)$, i.e., coordinates in which the symplectic form ω equals $\sum_{i=1}^n dq_i \wedge dp_i$, the Poisson bracket is of the form

$$\boxed{\{f, g\} = \textstyle\sum_{i=1}^n \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right)}$$

due to (10.1.1); in particular

$$\{p_i, p_k\} = \{q_i, q_k\} = 0 \quad \text{and} \quad \{q_i, p_k\} = \delta_{i,k} \qquad (i, k = 1, \ldots, n). \qquad \Diamond$$

As the Poisson bracket $\{f, g\}$ again defines a Hamiltonian vector field $X_{\{f,g\}}$, it is natural to ask how this vector field relates to X_f and X_g. This leads us to the notion of a Lie bracket.

In (A.3.3), when writing a vector field X in local coordinates $(z_1, \ldots, z_n)$ as

$$X(z) = \sum_{i=1}^n X_i(z) \frac{\partial}{\partial z_i}\,,$$

we used the isomorphism between vector fields and first order differential operators as given by the Lie derivative of functions. Then obviously a composition of Lie derivatives $L_X L_Y$ is a second order differential operator. But interestingly enough, the following lemma holds:

10.19 Lemma *The operator $L_X L_Y - L_Y L_X$ is a first order differential operator.*

Proof In local coordinates $(z_1, \dots, z_n)$,

$$L_X L_Y \varphi = \sum_{i=1}^n X_i \frac{\partial}{\partial z_i}\left(\sum_{j=1}^n Y_j \frac{\partial \varphi}{\partial z_j}\right) = \sum_{i=1}^n \sum_{j=1}^n \left(X_i \frac{\partial Y_j}{\partial z_i}\frac{\partial \varphi}{\partial z_j} + X_i Y_j \frac{\partial^2 \varphi}{\partial z_i \partial z_j}\right),$$

hence

$$(L_X L_Y - L_Y L_X)\varphi = \sum_{i=1}^n \sum_{j=1}^n \left(X_i \frac{\partial Y_j}{\partial z_i} - Y_i \frac{\partial X_j}{\partial z_i}\right)\frac{\partial \varphi}{\partial z_j}.$$

□

We can therefore define, in view of the above-mentioned isomorphism:

10.20 Definition *The* **Lie bracket** *or* **commutator** *of two vector fields $X, Y : M \to TM$, denoted as $[X, Y]$, is the vector field on M that satisfies*

$$L_{[X,Y]} = L_X L_Y - L_Y L_X .$$

In local coordinates $(z_1, \dots, z_n)$, our lemma implies that

$$[X, Y] = \sum_{j=1}^n \sum_{i=1}^n \left(X_i \frac{\partial Y_j}{\partial z_i} - Y_i \frac{\partial X_j}{\partial z_i}\right)\frac{\partial}{\partial z_j}.$$

10.21 Theorem *The flows Φ_t and Ψ_s generated by complete vector fields X and Y commute if and only if $[X, Y] = 0$, see the figure.*

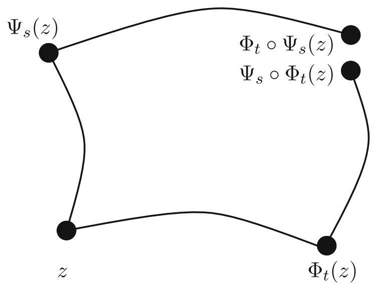
Proof

- If $\Phi_t \circ \Psi_s(z) = \Psi_s \circ \Phi_t(z)$

  for all $s, t \in \mathbb{R}$, then for all functions $f \in C^\infty(M, \mathbb{R})$,

  $$L_{[X,Y]} f = L_X L_Y f - L_Y L_X f = \left.\left(\frac{d}{dt}\Phi_t^* \frac{d}{ds}\Psi_s^* f - \frac{d}{ds}\Psi_s^* \frac{d}{dt}\Phi_t^* f\right)\right|_{t=s=0} = 0.$$

- The proof of the converse can be found for example in GIAQUINTA and HILDEBRANDT [GiHi], § 9.1.4. □



Non-commuting flows Φ and Ψ

10.22 Theorem

For functions $f, g \in C^\infty(M, \mathbb{R})$ on a symplectic manifold (M, ω), one has

$$d\{f, g\} = -\mathbf{i}_{[X_f, X_g]}\omega\,.$$

10.23 Remark Therefore the commutator of the Hamiltonian vector fields of f and g is again a Hamiltonian vector field, whose Hamilton function is $-\{f, g\}$:

$$X_{\{f,g\}} = -[X_f, X_g]\,. \qquad \Diamond$$

10.24 Lemma *For vector fields $X, Y : M \to TM$ and k-forms α on M, one has*

$$\mathbf{i}_{[X,Y]}\alpha = L_X \mathbf{i}_Y \alpha - \mathbf{i}_Y L_X \alpha\,.$$

Proof

- For 0-forms, the relation is trivial, because the inner product of a function and a vector field vanishes.
- For a local coordinate system $(z_1, \ldots, z_n)$ on M and the 1-forms $\alpha := dz_k$, one gets

$$L_X \mathbf{i}_Y dz_k - \mathbf{i}_Y L_X dz_k = (L_X L_Y z_k - L_X d \underbrace{\mathbf{i}_Y z_k}_{0}) - \mathbf{i}_Y d L_X z_k$$
$$= L_X L_Y z_k - L_Y L_X z_k + d \underbrace{\mathbf{i}_Y L_X z_k}_{0} = L_{[X,Y]} z_k = \underbrace{d\mathbf{i}_{[X,Y]} z_k}_{0} + \mathbf{i}_{[X,Y]} dz_k\,.$$

- More general forms are combined from functions and the 1-forms dz_k by exterior products, and one uses (B.5.2). □

Proof of Theorem 10.22:
In view of $d\omega = 0$, one calculates

$$d\{f, g\} = d\mathbf{i}_{X_g}\mathbf{i}_{X_f}\omega = L_{X_g}\mathbf{i}_{X_f}\omega - \mathbf{i}_{X_g} d\mathbf{i}_{X_f}\omega$$
$$= L_{X_g}\mathbf{i}_{X_f}\omega - \mathbf{i}_{X_g} L_{X_f}\omega + \mathbf{i}_{X_g}\mathbf{i}_{X_f} d\omega = L_{X_g}\mathbf{i}_{X_f}\omega - \mathbf{i}_{X_f} L_{X_g}\omega = -\mathbf{i}_{[X_f, X_g]}\omega,$$

because $L_{X_f}\omega = L_{X_g}\omega = 0$. □
The fact that ω is closed has entered into the proof in an essential manner.

10.25 Theorem (Lie Algebra Properties of Poisson Bracket)
For $f, g, h \in C^\infty(M, \mathbb{R})$, one has the properties

$$\{f, g\} = -\{g, f\} \quad \textit{(Antisymmetry)}$$

and

$$\{\{f, g\}, h\} + \{\{g, h\}, f\} + \{\{h, f\}, g\} = 0 \quad \textit{(Jacobi Identity)}\,. \tag{10.2.2}$$

Proof

- The antisymmetry of the Poisson bracket follows from the antisymmetry of ω.
- We have $\{\{f, g\}, h\} = -L_{X_{\{f,g\}}} h$,
$$\{\{g, h\}, f\} = -L_{X_f} L_{X_g} h \quad \text{and} \quad \{\{h, f\}, g\} = +L_{X_g} L_{X_f} h\,,$$
hence the left side of the Jacobi identity equals
$$(L_{X_g} L_{X_f} - L_{X_f} L_{X_g})h - L_{X_{\{f,g\}}} h = (L_{[X_g, X_f]} + L_{X_{\{g,f\}}})h = 0\,. \qquad \square$$

10.26 Remark
Consequently, the $\mathbb{R}$-vector space $C^\infty(M, \mathbb{R})$ on a symplectic manifold (M, ω), together with the Poisson bracket, is a Lie algebra (see page 553). ◊

10.3 Canonical Transformations

10.27 Definition

- *Let (P, ω) and (Q, ρ) be symplectic manifolds with* $\dim(P) = \dim(Q)$. *A mapping*
$$F \in C^1(P, Q) \quad \text{satisfying} \quad F^*\rho = \omega$$
is called **symplectic**, *or a* **canonical transformation**.
- *A symplectic diffeomorphism $F \in C^1(P, Q)$ is called a* **symplectomorphism**.
- *If a symplectomorphism $F \in C^1(P, P)$ can be written as the solution $F = F_1$ of a differential equation $\frac{\mathrm{d}}{\mathrm{d}t} F_t = X_{H_t} \circ F_t$, $F_0 = \mathrm{Id}_P$ with explicit time dependence, and the Hamiltonian vector field X_{H_t} of a time dependent Hamiltonian $H_t : P \to \mathbb{R}$, it will be called a* **Hamiltonian** *symplectomorphism.*

10.28 Remark

1. Canonical transformations are the structure preserving mappings of symplectic manifolds, in a similar sense as linear mappings preserve the structure of vector spaces.
2. The symplectomorphisms of (P, ω) form a subgroup of the group of diffeomorphisms of P, and the Hamiltonian symplectomorphisms in turn form a subgroup of the group of symplectomorphisms.
3. The books [LiMa] by LIBERMANN and MARLE, by MCDUFF and SALAMON [MS], and by HOFER and ZEHNDER [HZ, Zeh] discuss further geometric and topological aspects. ♢

10.29 Examples (Symplectomorphisms)

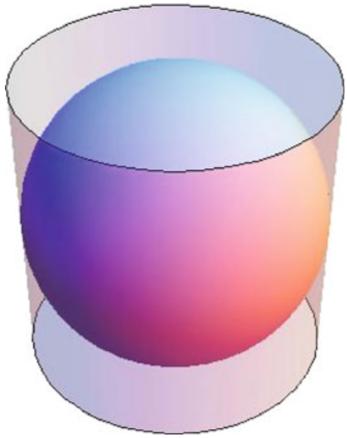
1. For the symplectic manifolds
$$(P, \omega) := \big((0,\infty)\times\mathbb{R}, r\,dr\wedge d\varphi\big) \quad\text{and}\quad (Q,\rho) := \big(\mathbb{R}^2, dx_1\wedge dx_2\big)\,,$$
the transformation into polar coordinates $F(r,\varphi) := (r\cos\varphi, r\sin\varphi)$ is a canonical transformation (but is neither injective nor surjective).
2. If $\Phi : \mathbb{R}\times P \to P$ is a Hamiltonian flow on the symplectic manifold (P,ω), then by Theorem 10.13, the $\Phi_t : P \to P$ are symplectomorphisms.
3. The translations $\Phi_a : \mathbb{T}^2 \to \mathbb{T}^2,\ x \mapsto x + a \quad (a \in \mathbb{T}^2)$ of the 2-torus $\mathbb{T}^2$ with the area form ω are symplectic, but only in the case $\Phi_0 = \mathrm{Id}_{\mathbb{T}^2}$ are they Hamiltonian:
For $F_1 = \Phi_a$, $F_0 = \mathrm{Id}_{\mathbb{T}^2}$, $\frac{\mathrm{d}}{\mathrm{d}t}F_t = X_{H_t}\circ F_t$ with a time dependent Hamiltonian vector field $t \mapsto X_{H_t}$ (and with the identification $T\mathbb{T}^2 \cong \mathbb{R}^2\times\mathbb{T}^2$), one has
$$\begin{aligned} a &= \int_{\mathbb{T}^2} a\,\omega = \int_{\mathbb{T}^2}\big(F_1(x)-x\big)\,\omega(x) = \int_{\mathbb{T}^2}\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}F_t(x)\,\mathrm{d}t\,\omega(x)\\ &= \int_{\mathbb{T}^2}\int_0^1 X_{H_t}\circ F_t(x)\,\mathrm{d}t\,\omega(x) = \int_0^1\int_{\mathbb{T}^2} X_{H_t}\circ F_t(x)\,\omega(x)\,\mathrm{d}t\\ &= \int_0^1\int_{\mathbb{T}^2} X_{H_t}(x)\,\omega(x)\,\mathrm{d}t = \mathbb{J}\int_0^1\int_{\mathbb{T}^2}\nabla H_t(x)\,\omega(x)\,\mathrm{d}t = 0\,. \end{aligned}$$
Here the inner integral vanishes according to Stokes' theorem (Theorem B.39). ♢

10.30 Exercise (Sphere and Cylinder)[2] Show that for the cylinder

$$\mathcal{Z} := \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 = 1,\ |x_3| < 1\},$$

the radial projection onto the sphere, i.e., the mapping $F : \mathcal{Z} \to S^2$,

$$x \mapsto \left(x_1\sqrt{1-x_3^2},\ x_2\sqrt{1-x_3^2},\ x_3\right)^\top,$$



is a symplectomorphism onto its image (i.e., onto the sphere minus its poles). Hereby, the underlying symplectic forms are the area forms of the surfaces imbedded into $\mathbb{R}^3$. For instance, for the sphere, this is

$$\omega_x(Y, Z) := \det(x, Y, Z) \quad \left(Y, Z \in T_x S^2\right). \Diamond$$

Locally, in any symplectic manifold, one can always use the canonical coordinates from Remark 10.18. In other words, locally, all symplectic manifolds of a given dimension look alike:

10.31 Theorem (Darboux) *For every point $x \in P$ of a $2n$-dimensional symplectic manifold (P, ω), there exists a chart (U, φ) near x with coordinates $\varphi = (p_1, \ldots, p_n, q_1, \ldots, q_n)$, such that $\omega\restriction_U = \sum_{i=1}^n dq_i \wedge dp_i$.*

Proof

• We work in the images of local charts at x. So we show that for any two open neighborhoods U_0, U_1 of $0 \in \mathbb{R}^{2n}$ and symplectic forms ω_k on U_k, there exist neighborhoods $V_k \subseteq U_k$ of 0 and a diffeomorphism $F : V_0 \to V_1$ such that $\omega_0\restriction_{V_0} = F^*\left(\omega_1\restriction_{V_1}\right)$.

In each step of the proof, we may reduce the neighborhoods U_k as needed, without renaming them.

• We assume that by means of a linear map, it has already been achieved that $\omega_0(0) = \omega_1(0)$. The linear Darboux theorem (Theorem 6.13) shows that this is possible.

• The diffeomorphism will be constructed as solution $F = F_1$ to a time dependent differential equation $\frac{\mathrm{d}}{\mathrm{d}t} F_t = X_t \circ F_t$ with initial value $F_0 = \mathrm{Id}$. The diffeomorphisms

[2]This is actually the title of an opus by Archimedes published in 225 BC. In it, he proved among other things that a sphere and its circumscribed cylinder have the same area.
When Cicero was questor on Sicily he looked for the tomb of Archimedes, who had been murdered, and believed to have found it. “Above the brush, I noticed, there rose a small column, on which there were the shapes of a sphere and a cylinder.” (45 BC, *Tusculanae disputationes*).

F_t will satisfy $F_t(0) = 0$ and $\mathrm{D}F_t(0) = 1\!\!1$. The vector field X_t will be chosen to satisfy the condition

$$L_{X_t}\omega_t = \omega_0 - \omega_1 \quad \text{with} \quad \omega_t := (1-t)\omega_0 + t\omega_1 \tag{10.3.1}$$

(The last bullet point explains why this is possible.) This implies that $F_t^*\omega_t = \omega_0$, because $F_0^*\omega_0 = \omega_0$ and

$$\tfrac{\mathrm{d}}{\mathrm{d}t} F_t^*\omega_t = F_t^*\left(L_{X_t}\omega_t + \tfrac{\mathrm{d}}{\mathrm{d}t}\omega_t\right) = F_t^*(0) = 0\,.$$

So F_t deforms the symplectic form.
• On a small neighborhood of zero, ω_t is a symplectic form for all $t \in [0, 1]$, because $\omega_t(0)$ is constant and non-degenerate, and to be non-degenerate is an open condition.
• By the Poincaré lemma (see Appendix B.7), the closed form $\omega_1 - \omega_0$ is exact in some ball around 0, so it is equal to $\mathrm{d}\theta$ for an appropriate 1-form θ. As ω_t in (10.3.1) is closed, its Lie derivative is $L_{X_t}\omega_t = d\mathbf{i}_{X_t}\omega_t$, so the requirement is that $\mathbf{i}_{X_t}\omega_t = \theta$ (possibly plus some df). Since ω_t is non-degenerate, this requirement can be met. By adding an appropriate df to θ, we normalize $\theta(0) = 0$, hence $X_t(0) = 0$, and thus $F_t(0) = 0$. □

In the remainder of the chapter, for simplicity, we will restrict our discussion to symplectic manifolds that are cotangent bundles $(P, \omega) = (T^*M, \omega_0)$ of a manifold M (with the canonical symplectic form $\omega_0 = -d\theta_0$ from Definition 10.7). They have a particular class of canonical transformations:

10.32 Definition *For a diffeomorphism $f \in C^1(M, N)$, the mapping*

$$T^*f : T^*N \to T^*M\,,\ \left\langle T^*f(\beta_n), v\right\rangle := \langle \beta_n, Tf(v)\rangle \quad \left(\beta_n \in T_n^*N\,,\ v \in T_mM\right)$$

with $n := f(m)$ is called the **cotangent map** *or the* **cotangent lift** *of f.*

10.33 Remark

1. As the derivatives of f at $m \in M$, namely the linear mappings $T_mf : T_mM \to T_nN$, are surjective, the mappings

$$T_n^*f := T^*f\restriction_{T_n^*N} : T_n^*N \to T_m^*M \qquad (n \in f(M) \subseteq N)$$

   are isomorphisms of the cotangent spaces. As f itself is surjective, T_n^*f is defined for all $n \in N$.
2. So this is the dual notion to the tangent map $Tf : TM \to TN$ introduced in Definition A.47. The diagram to the right, with the base point projection π_M^*

introduced in (10.1.2), commutes. Note however that the tangent map (in contradistinction to the cotangent map) is also defined for mappings $f \in C^1(M, N)$ that are not diffeomorphisms.

$$\begin{array}{ccc} T^*N & \xrightarrow{T^*f} & T^*M \\ \pi_N^* \downarrow & & \downarrow \pi_M^* \\ N & \xrightarrow{f^{-1}} & M \end{array}$$

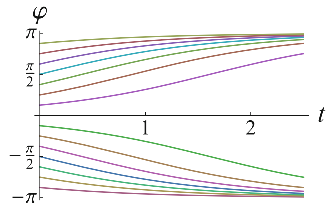
Moreover, for diffeomorphisms $g : L \to M$ and $f : M \to N$, one has

$$T(f \circ g) = Tf \circ Tg \quad \text{, but} \quad T^*(f \circ g) = T^*g \circ T^*f \,. \tag{10.3.2}$$

3. In classical mechanics, these mappings are also called *point transformations*. ♢

10.34 Example (Cotangent Lift of a Gradient Flow) On the circle $S^1 = \left\{ \left(\begin{smallmatrix} \sin\varphi \\ \cos\varphi \end{smallmatrix}\right) \mid \varphi \in [-\pi, \pi] \right\}$, the height function $h : S^1 \to \mathbb{R}$, $\left(\begin{smallmatrix} \sin\varphi \\ \cos\varphi \end{smallmatrix}\right) \mapsto \cos\varphi$ generates the gradient vector field that is tangential to S^1 (see page 97),

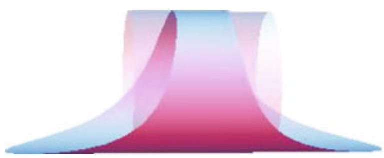
$$\nabla h \left(\left(\begin{smallmatrix} \sin\varphi \\ \cos\varphi \end{smallmatrix}\right) \right) = \sin\varphi \left(\begin{smallmatrix} -\cos\varphi \\ \sin\varphi \end{smallmatrix}\right) .$$



The gradient flow $\dot{x} = \nabla h(x)$ (intuitively speaking opposed to the direction of gravity!) therefore corresponds to the differential equation $\dot{\varphi} = -\sin\varphi$, which has

$$f_t(\varphi) = 2 \operatorname{arccot}\big(e^t \cot(\varphi/2)\big) \quad (t \in \mathbb{R})$$

as a solution (see the figure).
The cotangent lift $T^* f_t(p, \varphi) = \big((\cosh(t) + \sinh(t)\cos(\varphi))\, p\,,\ f_{-t}(\varphi)\big)$ of the diffeomorphism f_t is an area preserving mapping of the cylinder $\mathbb{R} \times S^1 \cong T^*S^1$.
In the figure on the right, one can see the (transparent) cylinder $[-1, 1] \times S^1$ and its image under the cotangent lift $T^* f_1$. ♢



10.35 Theorem (Cotangent Lift) *The cotangent lift T^*f of a diffeomorphism $f \in C^1(M, M)$ leaves the tautological 1-form θ_0 on T^*M invariant, i.e.,*

$$(T^*f)^*\, \theta_0 = \theta_0 \,.$$

Proof In local vector bundle coordinates (p, q) and (P, Q) with $Q := f(q)$, one has $p = \mathrm{D}f(q)^\top P$, hence $P = \big(\mathrm{D}f(q)^\top\big)^{-1} p$ and

$$\theta_0 = \langle P, dQ\rangle = \left\langle \left(\mathrm{D}f(q)^\top\right)^{-1} p, \mathrm{D}f(q)\, dq\right\rangle = \langle p, dq\rangle = (T^* f)^*\theta_0 \,.$$

Here we have used the abbreviated notation $\langle P, dQ\rangle := \sum_{i=1}^{\dim M} P_i\, dQ_i$. □

Thus in particular, cotangent lifts are symplectic.

Another class of symplectic mappings is given by certain fiber translations:

10.36 Definition *A* **fiber translation** *of a cotangent bundle T^*M is a mapping*

$$\mathrm{trans}_A : T^*M \to T^*M \quad , \quad \mathrm{trans}_A(p) = p + A(q) \quad \left(q \in M,\ p \in T_q^*M\right),$$

with a 1-form $A \in \Omega^1(M)$.

10.37 Theorem *A fiber translation* trans_A *is symplectic if and only if A is closed ($dA = 0$), and it is Hamiltonian if A is exact ($A = dh$).*

Proof

- In local canonical coordinates (p, q), the symplectic form $\sum_k dq_k \wedge dp_k$ is transformed into $\sum_k dq_k \wedge d\big(p_k - A_k(q)\big)$ by pull-back with the fiber translation of $A(q) = \sum_k A_k(q) dq_k$. So it changes by dA.

- If $A = dh$, then the lift $H = h \circ \pi_M^* : T^*M \to \mathbb{R}$ of $h : M \to \mathbb{R}$ generates a Hamiltonian flow $\Phi : \mathbb{R} \times T^*M \to T^*M$ with $\Phi_{-1} = \mathrm{trans}_A$, because the Hamiltonian vector field X_H equals $-\sum_k \frac{\partial h}{\partial q_k}(q)\frac{\partial}{\partial p_k} = -\sum_k A_k(q)\frac{\partial}{\partial p_k}$ in local coordinates, hence $\Phi_t(p, q) = \big(p - tA(q), q\big)$. □

As with all transformations, we can take an active or a passive point of view for canonical transformations. In the first case, we are interested in how phase space points are mapped by F; in the second, how a a coordinate system (like for instance the coordinates $(p_1, \ldots, p_n, q_1, \ldots, q_n)$ on the vector space $\mathbb{R}_p^n \times \mathbb{R}_q^n$) becomes a new coordinate system under F^*.

An important motivation for using canonical transformations is to solve the Hamiltonian differential equations

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} \quad , \quad \dot{q}_i = \frac{\partial H}{\partial p_i} \qquad (i = 1, \ldots, n).$$

Namely, it is possible that in a new coordinate system (P, Q), the differential equations have a simpler form, so that we can solve them. It transpires that it is expedient in this process to choose not some arbitrary new coordinates $X_1, \ldots, X_{2n}$, but rather such as preserve the form of the Hamiltonian equation. This is why *canonical* transformations are recommended for a change of coordinates.

For a given Hamilton function $H : P := T^*M \to \mathbb{R}$, we consider, using the canonical 1-form θ_0 on P introduced in Definition 10.7, the 1-form

$$\Theta_H := \pi_1^*\theta_0 - H\, dt$$

on the *extended phase space* $P \times \mathbb{R}_t$. Here $\mathbb{R}_t$ represents the time axis, and $\pi_1 : P \times \mathbb{R}_t \to P$ is the projection on the first factor P.

Because $d\theta_0 = -\omega_0$, we conclude

$$d\Theta_H = -\pi_1^*\omega_0 - dH \wedge dt\,.$$

This 2-form must be degenerate as it is an antisymmetric bilinear form on the odd-dimensional space $P \times \mathbb{R}_t$. In particular, the vector field $\tilde{X}_H$ on $P \times \mathbb{R}_t$ given by

$$\tilde{X}_H(x,t) := X_H(x) + \frac{\partial}{\partial t} \tag{10.3.3}$$

satisfies $\mathbf{i}_{\tilde{X}_H} d\Theta_H = -dH + dH = 0$. On the other hand:

10.38 Lemma *If a vector field W on $P \times \mathbb{R}_t$ satisfies*

$$d\Theta_H(W,\cdot) \equiv 0\,,$$

then

$$W = f\,\tilde{X}_H$$

for an appropriate function f on $P \times \mathbb{R}_t$ and the vector field $\tilde{X}_H$ from (10.3.3).

Proof The local form of a vector field on $P \times \mathbb{R}_t$ is

$$W = \sum_{i=1}^{n} \left(a_i \frac{\partial}{\partial q_i} + b_i \frac{\partial}{\partial p_i} \right) + c\frac{\partial}{\partial t}\,.$$

So the 1-form $d\Theta_H(W,\cdot)$ equals

$$\textstyle\sum_{i=1}^n \left[-a_i dp_i + b_i dq_i - \left(\frac{\partial H}{\partial p_i} b_i + \frac{\partial H}{\partial q_i} a_i \right) dt + c \left(\frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i \right) \right].$$

If this vanishes, one gets by comparing coefficients that $a_i = +c\,\frac{\partial H}{\partial p_i}$ and $b_i = -c\,\frac{\partial H}{\partial q_i}$ $(i = 1,\ldots,n)$, hence $W = f\,\tilde{X}_H$ with $f = c$. □

So at each point x of $P \times \mathbb{R}_t$, there exists in the local tangent space $T_x(P \times \mathbb{R}_t)$ a subspace of dimension exactly 1, consisting of vectors that produce a vanishing 1-form when plugged into $d\Theta_H$.

Following these directions, we obtain curves $c : I \to P \times \mathbb{R}_t$, which we can even choose to be parametrized by t; these curves are called *characteristic* or *vortex lines*

$$c(t) = \big(p_1(t),\ldots,p_n(t),\ q_1(t),\ldots,q_n(t),\ t\big) \qquad (t \in I).$$

Then it follows by Lemma 10.38 that $\dot{p}_i = -\frac{\partial H}{\partial q_i}$ and $\dot{q}_i = \frac{\partial H}{\partial p_i}$, namely the Hamiltonian differential equations.

10.39 Theorem *If the coordinate change* $(p,q) \mapsto \big(P(p,q), Q(p,q)\big)$ *on the phase space* $M \subseteq \mathbb{R}^n_p \times \mathbb{R}^n_q$ *gives rise to a canonical transformation* $g : M \to \mathbb{R}^n_P \times \mathbb{R}^n_Q$*, then the Hamiltonian differential equations* $\dot{p}_i = -\frac{\partial H}{\partial q_i}$, $\dot{q}_i = \frac{\partial H}{\partial p_i}$ *of* $H : M \to \mathbb{R}$ *are transformed into*

$$\dot{P}_i = -\frac{\partial K}{\partial Q_i} \quad , \quad \dot{Q}_i = \frac{\partial K}{\partial P_i} \quad , \text{with} \quad K\big(P(p,q), Q(p,q)\big) = H(p,q)\,.$$

Proof We consider the 1-form $\alpha := \sum_{i=1}^n (p_i dq_i - P_i dQ_i)$ on the coordinate chart in M. Since g is canonical, we have $d\alpha = 0$. Therefore on the extended phase space, it follows that

$$\pi_1^* \sum_{i=1}^n p_i dq_i - H dt = \pi_1^* \sum_{i=1}^n P_i dQ_i - H dt + \pi_1^* \alpha\,.$$

If we subtract $\pi_1^*\alpha$ on the right hand side, the vortex lines remain the same, because they only depend on the exterior derivative, and $d\pi_1^*\alpha = \pi_1^* d\alpha = 0$. As the vortex lines are the same, so is the Hamiltonian equation. □

It is not just the Hamilton function generating the equations of motion that transforms in a simple manner under canonical transformations, but also the Poisson brackets:

10.40 Theorem *Assume the diffeomorphism* $F : P \to Q$ *is a canonical transformation of the symplectic manifolds* (P,ω) *and* (Q,ρ)*. Then, for* $f, g \in C^\infty(Q,\mathbb{R})$*, the pull-back*[3] F^* *satisfies*

$$F^* X_f = X_{F^* f} \quad \text{and} \quad F^*\big(\{f,g\}_Q\big) = \{F^* f, F^* g\}_P\,.$$

Proof

- The first identity follows from

$$\mathbf{i}_{F^* X_f}\omega = \mathbf{i}_{F^* X_f} F^*\rho = F^*(\mathbf{i}_{X_f}\rho) = F^*(df) = d(F^* f) = \mathbf{i}_{X_{F^* f}}\omega\,.$$

  because ω is nondegenerate.

- The second identity follows from the first one: $F^*\big(\{f,g\}_Q\big) = F^*\big(\mathbf{i}_{X_g}\mathbf{i}_{X_f}\rho\big) =$

$$= \mathbf{i}_{F^* X_g}\mathbf{i}_{F^* X_f} F^*\rho = \mathbf{i}_{X_{F^* g}}\mathbf{i}_{X_{F^* f}}\omega = \{F^* f, F^* g\}_P\,.$$ □

---

[3]**Definition:** The *pull-back* $F^* X$ of a vector field $X : Q \to TQ$ on a manifold Q with respect to a diffeomorphism $F : P \to Q$ is the vector field $F^* X := T(F^{-1}) \circ X \circ F$ on P.

10.4 Lagrangian Manifolds

"Symplectic Creed: Everything is a Lagrange manifold."
ALAN WEINSTEIN, in [Wein]

The Hamiltonian equations of motion are distinguished from other systems of ordinary differential equations by the fact that the information incorporated in them is encoded in a single function, the Hamilton function. Similarly (albeit subject to certain restrictions), canonical transformations can be represented by a single function, called the *generating* function. To understand this representation, we introduce the notion of a Lagrangian manifold.

Let us first recall the definition of Lagrangian subspaces $L \subset E$ of a symplectic vector space (E, ω) in Chapter 6.4: they are isotropic (i.e., ω vanishes on L) and of maximal dimension ($\dim(L) = \frac{1}{2}\dim(E)$). This notion generalizes easily from symplectic vector spaces to symplectic manifolds:

10.41 Definition *Let (P, ω) be a symplectic manifold and $I : L \to P$ the imbedding of a submanifold L.*[4]
L is called **isotropic** if $I^*\omega = 0$, and **Lagrangian** if moreover $\dim L = \frac{1}{2}\dim P$.

10.42 Examples (Lagrangian Manifolds)

1. For $\dim(P) = 2$, every 1-dimensional submanifold L is Lagrangian, because $I^*\omega$ is a 2-form on L and therefore $I^*\omega = 0$.
2. Let $\dim(P) = 2n$ and $F_1, \ldots, F_n \in C^\infty(P, \mathbb{R})$. Let $f \in F(P)$ be a regular value of
$$F := \begin{pmatrix} F_1 \\ \vdots \\ F_n \end{pmatrix} : P \to \mathbb{R}^n .$$
Then $L := F^{-1}(f)$ is an n-dimensional submanifold of P.

   If $\{F_i, F_k\} = 0$ for $i, k \in \{1, \ldots, n\}$, then L is Lagrangian.
   Proof: By regularity of F, the n-form $dF_1 \wedge \ldots \wedge dF_n\ (x) \neq 0$ for any $x \in L$; and because of the relation $\mathbf{i}_{X_{F_i}}\omega = dF_i$, the Hamiltonian vector fields $X_{F_1}, \ldots, X_{F_n}$ are linearly dependent at each $x \in L$. Moreover, they are tangential to L because $dF_i(X_{F_k}) = \mathbf{i}_{X_{F_k}} dF_i = \mathbf{i}_{X_{F_k}}\mathbf{i}_{X_{F_i}}\omega = \{F_i, F_k\} = 0$.

   As $\dim(T_x L) = n$, the vectors $X_{F_1}(x), \ldots, X_{F_n}(x)$ span the tangent space $T_x L$ of L at x. Hence, tangent vector fields Y, Z to L can be written as linear combinations $Y = \sum_{i=1}^n Y_i \cdot X_{F_i}$ with functions $Y_i : L \to \mathbb{R}$, and accordingly also for Z.

   Therefore $\omega(Y, Z) = 0$, hence $I^*\omega = 0$, and this means that L is isotropic, and because $\dim(L) = n$, Lagrangian.

[4]Submanifolds of manifolds P are defined in analogy to the case $P = \mathbb{R}^n$ (Definition 2.34).

3. Let M be a n-dimensional manifoldand $P := T^*M$ its cotangent bundle. Then on P, there exist the canonical 1-form θ_0 introduced in Definition 10.7, and the canonical symplectic form $\omega_0 = -d\theta_0$.
   In the example $M = \mathbb{R}^n_q$, one has $\theta_0 = \sum_{i=1}^n p_i\, dq_i$ and $\omega_0 = \sum_{i=1}^n dq_i \wedge dp_i$.

   We now consider a 1-form α on M.

   The graph L of α is an n-dimensional submanifold $L \subset P$ (see Figure 10.4.1).

   Interpreting the 1-form α as a mapping $\hat{\alpha} : M \to L \subset P$, we can therefore pull back the canonical 1-form θ_0 by this mapping $\hat{\alpha}$, and we obtain

$$\hat{\alpha}^*\theta_0 = \alpha\,, \tag{10.4.1}$$

   because by Definition (10.1.3) of θ_0, and using the base point projection $\pi^*_M : T^*M \to M$, we have

$$\big\langle \theta_0(\hat{\alpha}(q))\,,\,\zeta \big\rangle \;=\; \big\langle \alpha(q)\,,\, T\pi^*_M(\zeta)\big\rangle \qquad \big(\zeta \in T_{\hat{\alpha}(q)}P\big),$$

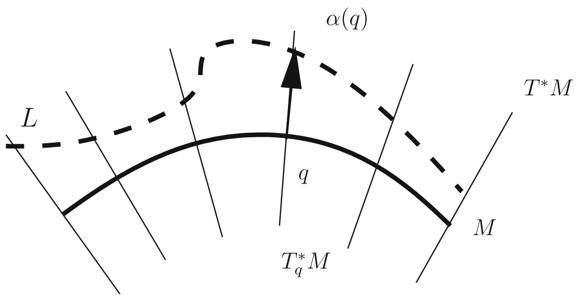


Figure 10.4.1 Lagrangian submanifold as the graph of a closed 1-form

   and therefore, with $\pi^*_M \circ \hat{\alpha} = \mathrm{Id}_M$, we obtain for $v_q \in T_qM$:

$$\begin{aligned}\big\langle \hat{\alpha}^*\theta_0(q)\,,\, v_q\big\rangle &= \big\langle \theta_0(\hat{\alpha}(q))\,,\, T(\hat{\alpha})(v_q)\big\rangle = \big\langle \alpha(q)\,,\, T(\pi^*_M)\circ T(\hat{\alpha})(v_q)\big\rangle \\ &= \big\langle \alpha(q), T(\pi^*_M \circ \hat{\alpha})(v_q)\big\rangle = \big\langle \alpha(q), T\mathrm{Id}_M(v_q)\big\rangle = \big\langle \alpha(q), v_q\big\rangle.\end{aligned}$$

   The exterior derivative of (10.4.1) is

$$d\alpha = d\hat{\alpha}^*\theta_0 = \hat{\alpha}^* d\theta_0 = -\hat{\alpha}^*\omega_0\,,$$

   and therefore $L = \mathrm{graph}(\alpha)$ is Lagrangian if and only if α is closed.[5] ◊

[5] Compare this with Theorem 10.37 on fiber translations.

In complete analogy to the linear case (Theorem 6.49), one obtains the following statement:

10.43 Theorem *Let $F : M_1 \to M_2$ be a diffeomorphism between the symplectic manifolds (M_i, ω_i). Then F is symplectic if and only if the graph*

$$\Gamma_F := \big\{\big(x, F(x)\big) \mid x \in M_1\big\} \subset M_1 \times M_2$$

of F is Lagrangian with respect to the symplectic form $\omega_1 \ominus \omega_2$ on $M_1 \times M_2$.

Proof The tangent space $T_{(x,F(x))}\Gamma_F$ of Γ_F at point $\big(x, F(x)\big)$ is of the form

$$T_{(x,F(x))}\Gamma_F = \big\{\big(v, TF(v)\big) \mid v \in T_x M_1\big\}.$$

Therefore, the imbedding $I : \Gamma_F \to M_1 \times M_2$ and $\omega := \omega_1 \ominus \omega_2$ (see (6.48)) satisfy

$$\begin{aligned} &(I^*\omega)\,\big((v_1, TF(v_1))\,,\ (v_2, TF(v_2))\big) \\ &\quad = \omega_1(v_1, v_2) - \omega_2\big(TF(v_1), TF(v_2)\big) = \big(\omega_1 - F^*\omega_2\big)(v_1, v_2)\,. \end{aligned}$$ □

10.5 Generating Functions of Canonical Transformations

We will now show that canonical transformations can, at least locally, be represented in terms of a single function, called a *generating function* of the canonical transformation.

We know by Theorem 10.43 that the graphs of canonical transformations are Lagrangian submanifolds.

So let now (M, ω) be a symplectic manifold and $I : L \to M$ a Lagrangian submanifold. By the Poincaré Lemma (see Appendix B.7), every $x \in M$ has a neighborhood $U \subset M$ of x and a 1-form θ on U with $\omega\restriction_U = -d\theta$. Now if $x \in L$, then $I^*\theta$ is closed because

$$0 = I^*\omega\restriction_U = -I^*d\theta = -dI^*\theta\,.$$

Therefore, again by the Poincaré Lemma, on an appropriate neighborhood $V \subset L$ of x, there exists a function $S : V \to \mathbb{R}$ such that

$$-I^*\theta\restriction_V = dS\,.$$

Such a function will be called a *generating function for L*.

Let us now consider in particular a canonical transformation F from (M_1, ω_1) to (M_2, ω_2), where these symplectic manifolds are assumed to be exact symplectic, i.e., $\omega_i = -d\theta_i$. Then the symplectic manifold (M, ω),

$$M := M_1 \times M_2 \quad \text{and} \quad \omega := \omega_1 \ominus \omega_2 \,,$$

is also exact symplectic:

$$\omega = -d\theta \quad \text{with} \quad \theta := \theta_1 \ominus \theta_2 = \pi_1^*\theta_1 - \pi_2^*\theta_2 \quad \text{on} \quad M.$$

By Theorem 10.43, the graph $\Gamma_F \subset M$ is a Lagrangian submanifold. At least locally (namely by restricting F to an appropriate neighborhood $U_1 \subset M_1$ of $m_1 \in M_1$), we can find a generating function S such that

$$I^*\theta = -dS \,.$$

We use local canonical coordinates $(P, Q) = (P_1, \ldots, P_n, Q_1, \ldots, Q_n)$ on U_1 and $(p, q) = (p_1, \ldots, p_n, q_1, \ldots, q_n)$ on $U_2 := F(U_1) \subseteq M_2$, where

$$F(P, Q) = (p, q)\,.$$

By Darboux' theorem (Theorem 6.13), we may assume $\theta_2 = \sum_{i=1}^n p_i\, dq_i$ and $\theta_1 = \sum_{i=1}^n P_i\, dQ_i$ on U_1 and U_2 respectively, so that

$$\theta = \sum_i \bigl(P_i\, dQ_i - p_i\, dq_i\bigr)\,.$$

For $n = 1$, the canonical form can be written in at least two of the following four forms, because at least two entries in $\mathrm{D}F(P, Q) \in \mathrm{Mat}(2, \mathbb{R})$ do not vanish:

1. We write S as a function $S_1(q, Q)$ of $(q_1, \ldots, q_n, Q_1, \ldots, Q_n)$. From $dS_1 = -I^*\theta_1$, it follows that $dS_1 = \sum_i (\frac{\partial S_1}{\partial q_i} dq_i + \frac{\partial S_1}{\partial Q_i} dQ_i) = \sum_i (p_i dq_i - P_i dQ_i)$, hence
$$p_i = \frac{\partial S_1}{\partial q_i} \quad , \quad P_i = -\frac{\partial S_1}{\partial Q_i} \qquad (i = 1, \ldots, n).$$
2. If S is set up as a function $S_2(q, P)$, it is expedient to add an exact form to θ_1: $dS_2 = -I^*\theta_2$ with $\theta_2 := \theta_1 - d\left(\sum_i Q_i P_i\right) = \sum_i (-Q_i dP_i - p_i dq_i)$, thus
$$Q_i = \frac{\partial S_2}{\partial P_i} \quad , \quad p_i = \frac{\partial S_2}{\partial q_i} \qquad (i = 1, \ldots, n).$$
3. If we set S as $S_3(p, Q)$, one obtains with $\theta_3 := \theta_1 + \sum_i d(p_i q_i) = \sum_i P_i dQ_i + q_i dp_i$ that
$$P_i = -\frac{\partial S_3}{\partial Q_i} \quad , \quad q_i = -\frac{\partial S_3}{\partial p_i} \qquad (i = 1, \ldots, n).$$
4. If we take $\theta_4 := \theta_1 + d\left(\sum -Q_i P_i + q_i p_i\right) = \sum_i -Q_i dP_i + q_i dp_i$, then S, expressed in terms of $S_4(p, P)$, $-dS_4 = I^*\theta_4$, yields

$$q_i = -\frac{\partial S_4}{\partial p_i} \quad , \quad Q_i = \frac{\partial S_4}{\partial P_i} \qquad (i = 1, \ldots, n).$$

In all four cases, we equally get $\omega = -d\theta_k \qquad (k = 1, 2, 3, 4)$.

Generally, $\dim(M_1 \times M_2) = 4n$. In this case, at least one of the coordinate $2n$-tuples can be used for the local representation of the canonical transformation. The argument can be found in Lemma 1 on 276 of Hofer and Zehnder, [HZ].

10.44 Example (Polar Coordinates and Harmonic Oscillator)
For $\omega > 0$ and

$$F : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^2 \quad , \quad (p, q) \mapsto (P, Q) = \left(\sqrt{\tfrac{p\omega}{\pi}} \cos(2\pi q), \sqrt{\tfrac{p}{\pi\omega}} \sin(2\pi q)\right),$$

one has

$$\begin{aligned} dQ \wedge dP &= \left(\tfrac{1}{2}\sqrt{\tfrac{1}{\pi\omega p}} \sin(2\pi q) dp + 2\sqrt{\tfrac{\pi p}{\omega}} \cos(2\pi q) dq\right) \wedge \\ &\quad \left(\tfrac{1}{2}\sqrt{\tfrac{\omega}{p\pi}} \cos(2\pi q) dp - 2\sqrt{p\omega\pi} \sin(2\pi q) dq\right) \\ &= dq \wedge dp \,, \end{aligned}$$

so the (local) diffeomorphism is area preserving.
For $S_1(q, Q) := -\frac{\omega}{2} Q^2 \cot(2\pi q)$, one gets

$$\frac{\partial S_1}{\partial q} = \frac{\pi\omega Q^2}{\sin^2(2\pi q)} \quad \text{and} \quad \frac{\partial S_1}{\partial Q} = -Q\omega \cot(2\pi q) \,.$$

Therefore S_1 generates the canonical transformation F:

$$\begin{aligned} \frac{P}{Q} &= \omega \cot(2\pi q) \text{ , hence } P = -\frac{\partial S_1}{\partial Q} \,, \\ \frac{Q^2}{\sin^2(2\pi q)} &= \frac{p}{\pi\omega} \text{ , hence } p = \frac{\partial S_1}{\partial q} \,. \end{aligned}$$

If we consider the Hamilton function $H(P, Q) := \frac{1}{2}(P^2 + \omega^2 Q^2)$, it gets transformed into

$$K(p, q) := \frac{\omega}{2\pi} p \,.$$

The linear equations of motion of H give rise to a constant vector field for K:

$$\dot{q} = \frac{\omega}{2\pi} \quad , \quad \dot{p} = 0 \,. \qquad \Diamond$$

10.45 Exercise (Representation of the Flow by Generating Functions)

(a) Show that for the Hamilton function of the harmonic oscillator

$$H_0 : \mathbb{R}^2 \to \mathbb{R} \quad , \quad H_0(p,q) = \tfrac{1}{2}(p^2 + q^2)$$

and times $t \in (-\pi/2, \pi/2)$, the solution (p_t, q_t) can be written in the form

$$p_t = p_0 - t\,\mathrm{D}_2 H_t(p_0, q_t) \quad , \quad q_t = q_0 + t\,\mathrm{D}_1 H_t(p_0, q_t)\,, \tag{10.5.1}$$

with the function

$$H_t(p,q) := \left[\frac{\sin(t)}{2t}\big(p^2 + q^2\big) + \frac{\cos(t) - 1}{t} pq\right] \Big/ \cos(t) \qquad \big(0 < |t| < \pi/2\big)$$

extending H_0.

(b) Generalize this result to the effect that for every quadratic Hamilton function $H_0 : \mathbb{R}^{2n} \to \mathbb{R}$, there exists a time $T > 0$ and generating functions H_t, $|t| < T$ such that a relation analogous to (10.5.1) holds. Calculate H_t from the solution (with $H_t(0) := 0$). Conclude that $(t, p, q) \mapsto H_t(p,q)$ is smooth also at $t = 0$.

(c) Show: Whereas the anharmonic oscillator with Hamilton function

$$H_0 : \mathbb{R}^2 \to \mathbb{R} \quad , \quad H_0(p,q) = \tfrac{1}{2}\big(p^2 + q^2 + q^4\big)$$

does define a dynamical system, a relation analogous to (10.5.1) cannot exist for any $t \neq 0$ on the entire phase space. ♢

Chapter 11
Motion in a Potential

Foucault pendulum. Image: courtesy of Miami University (Oxford, Ohio).
Photographer: Scott Kissell

This class of Hamiltonian motion is the most important one. It comprises both electrostatic and gravitational force fields.

It shares the property of reversibility with geodesic motion, and it can often be compared well with geodesic motion. Accordingly, many geometric techniques have been developed for the analysis of dynamics in a potential.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_11

11.1 Properties of General Validity

In this chapter, we study Hamiltonian systems on the phase space[1] $P := \mathbb{R}^d_p \times \mathbb{R}^d_q$ whose Hamiltonian has the form

$$H : P \to \mathbb{R} \quad , \quad H(p,q) = \tfrac{1}{2}\langle p, Ap\rangle + V(q)\,, \tag{11.1.1}$$

where $A = A^\top \in \mathrm{Mat}(d,\mathbb{R})$ is positive definite and $V \in C^2(\mathbb{R}^d,\mathbb{R})$ is a *potential*; such Hamiltonians occur frequently in physics. The Hamiltonian differential equation is therefore

$$\dot p = -\nabla V(q) \quad , \quad \dot q = Ap\,. \tag{11.1.2}$$

The *kinetic energy*, i.e., the quadratic form $K(p) := \frac{1}{2}\langle p, Ap\rangle$, can be diagonalized by a linear symplectic transformation $P \to P$, $(p,q) \mapsto (Op, Oq)$, where $O \in \mathrm{SO}(d)$ is a rotation. Denoting the reciprocals of the eigenvalues of A as $m_1,\ldots,m_d > 0$, one obtains $K(p) = \sum_{k=1}^d \frac{p_k^2}{2m_k}$. In terms of physics, these m_k will be interpreted as masses.

11.1.1 Existence of the Flow

For initial conditions $x_0 \in P$, the only way for the norm $\|\Phi_t(x)\|$ of the solution to become large is that, along with $\big(p(t), q(t)\big) := \Phi_t(x)$, the position $q(t)$ goes to infinity. This leads to the following sufficient condition for the existence of the flow.

11.1 Theorem *If there exists a constant $c > 0$ such that*

$$V(q) \ge -c\left(1 + \|q\|^2\right) \qquad \left(q \in \mathbb{R}^d\right), \tag{11.1.3}$$

then the differential equation (11.1.2) generates a flow $\Phi \in C^1(\mathbb{R} \times P, P)$.

Proof:
- For $m_{\min} := \min(m_1,\ldots,m_d)$, one has $\|\dot q\| \le \|p\|/m_{\min}$.

- Let $E := H(x_0)$. Then for all times $t \in (t_{\min}, t_{\max})$ from the maximal interval of existence, one has

$$\|\dot q(t)\| \le \frac{\|p(t)\|}{m_{\min}} \le c_1 \sqrt{\sum_{k=1}^d \tfrac{1}{2} p_k^2(t)/m_k} = c_1\sqrt{E - V(q(t))} \tag{11.1.4}$$

with $c_1 := \frac{\sqrt{2m_{\max}}}{m_{\min}}$. Letting $c_2 := c_1\sqrt{2\max(E,c)}$, we conclude from (11.1.3) that

[1] More generally, the configuration space is a manifold M and the phase space P is its cotangent bundle T^*M, see Chapter 10.1.

$$\|\dot{q}(t)\| \leq c_1\sqrt{E + c(1 + \|q(t)\|^2)} \leq c_2\sqrt{1 + \|q(t)\|^2}\,.$$

• Therefore, letting $f(t) := 1 + \|q(t)\|^2$ and $A := f(0)$, we conclude for all $t \in [0, t_{\max})$:

$$\begin{aligned} f(t) &= A + 2\int_0^t \langle q(s), \dot{q}(s)\rangle \,\mathrm{d}s \leq A + 2\int_0^t \|q(s)\|\,\|\dot{q}(s)\|\,\mathrm{d}s \\ &\leq A + 2c_2\int_0^t (1 + \|q(s)\|^2)\,\mathrm{d}s = A + 2c_2\int_0^t f(s)\,\mathrm{d}s\,. \end{aligned}$$

An analogous claim applies for $t \in (t_{\min}, 0]$. Then, from the Gronwall Lemma 3.42, we infer

$$f(t) \leq A\exp(2c_2|t|)\,,$$

so $f(t)$ is in particular bounded. Therefore, in view of (11.1.4) and (11.1.3), the solution $\Phi_t(x_0) = \big(p(t), q(t)\big)$ satisfies

$$\|\Phi_t(x_0)\|^2 = \|p(t)\|^2 + \|q(t)\|^2 \leq c_3\big(1 + \|q(t)\|^2\big) \leq c_3 A\exp(2c_2|t|)$$

with an appropriate constant c_3. By this estimate and local Lipschitz continuity of the right hand side of (11.1.2), the solution can be extended for all times.
• The continuous differentiability of the flow follows from the continuous differentiability of the right hand side of (11.1.2) and Theorem 3.45. □

11.2 Exercise (To Infinity in Finite Time)
Show that for $\varepsilon > 0$ and potential $V(q) := c\,(1 + \|q\|^2)^{1+\varepsilon}$, the differential equation (11.1.2) generates a complete flow if and only if $c \geq 0$. ♢

The kinetic energy K being nonnegative, for a total energy $E := H(p_0, q_0)$, the trajectory with initial condition (p_0, q_0) is confined to the connected component of q_0 of *Hill's Domain* $\{q \in \mathbb{R}^d \mid V(q) \leq E\}$.

11.1.2 Reversibility of the Flow

We now allow the configuration space M to be a manifold. On the symplectic cotangent bundle (T^*M, ω_0) instead of (11.1.1) we consider Hamiltonian functions $H \in C^2(P, \mathbb{R})$ with $P := T^*M$ that (in local bundle coordinates) have the form[2]

$$H(p, q) = \tfrac{1}{2}g_q^*(p, p) + V(q)\,. \tag{11.1.5}$$

[2]This includes open subsets $M \subseteq \mathbb{R}^d$ and $H(p, q) = \frac{\|p\|^2}{2m} + V(q)$ as a special case.

Here g denotes a Riemannian metric on M. So g_q is a positive definite bilinear form on the tangent space T_qM at q, and g_q^* its dual on T_q^*M, see also Example 8.4. If $V = 0$, then H generates (the cotangent form of) geodesic motion.

The motion generated by H is reversible:

11.3 Definition

- *The mapping* $\mathcal{T} : P \to P,\ (p, q) \mapsto (-p, q)$ *on the phase space* P *is called* **time reversal**.
- *A Hamiltonian* $H : P \to \mathbb{R}$ *(and the flow generated by it) is called* **reversible**, *if it is invariant under time reversal, i.e.,* $H \circ \mathcal{T} = H$.

$\mathcal{T}$ is an *antisymplectic* ($\mathcal{T}^*\omega_0 = -\omega_0$) diffeomorphism and an *involution*, i.e., $\mathcal{T} \circ \mathcal{T} = \mathrm{Id}_P$.

11.4 Examples (Reversibility)

1. The Hamiltonian (11.1.1) for motion in a potential is reversible.
2. For $B \in \mathbb{R} \setminus \{0\}$ and $\mathbb{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, the Hamiltonian

$$H : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R} \quad , \quad H(p, q) := \tfrac{1}{2}\|p - B\mathbb{J}q\|^2$$

from Section 6.3.3 is not reversible. It describes the motion of a charged particle in the plane under the influence of a constant magnetic field of strength B. $\diamondsuit$

11.5 Theorem *The flow* $\Phi : \mathbb{R} \times P \to P$ *generated by a reversible Hamiltonian* $H : P \to \mathbb{R}$ *of the form* (11.1.5) *satisfies*

$$\Phi_{-t} = \mathcal{T} \circ \Phi_t \circ \mathcal{T} \qquad (t \in \mathbb{R}). \tag{11.1.6}$$

Proof:

- As $\mathcal{T}$ is an involution, $(\mathcal{T} \circ \Phi_{t_1} \circ \mathcal{T}) \circ (\mathcal{T} \circ \Phi_{t_2} \circ \mathcal{T}) = \mathcal{T} \circ \Phi_{t_1+t_2} \circ \mathcal{T}$ and $\mathcal{T} \circ \Phi_0 \circ \mathcal{T} = \mathrm{Id}_P$. So both sides of (11.1.6) are dynamical systems.
  In order to prove their equality, it thus suffices to show that the time derivatives at time zero coincide.
- For the left hand side, we simply have $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{-t}^*\big|_{t=0} = -L_{X_H} = L_{-X_H}$.
  For the right hand side, $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{T}^*\Phi_t^*\mathcal{T}^*\big|_{t=0} = \mathcal{T}^*L_{X_H}\mathcal{T}^* = L_{\mathcal{T}^*X_H}$.
  But as $\mathcal{T}$ is antisymplectic ($\mathcal{T}^*\omega_0 = -\omega_0$) and H is reversible,

$$-\mathbf{i}_{\mathcal{T}^*X_H}\omega_0 = \mathcal{T}^*(\mathbf{i}_{X_H}\omega_0) = \mathcal{T}^*(dH) = d\mathcal{T}^*H = dH = \mathbf{i}_{X_H}\omega_0\,,$$

  so that $\mathcal{T}^*X_H = -X_H$, too. This proves (11.1.6). $\square$

If a Hamiltonian flow is reversible, and if there exists a time t_0 where the momentum $p(t_0) = 0$, then afterwards the particle traces back its trajectory:

$$\big(p(t_0 + s), q(t_0 + s)\big) = \big(-p(t_0 - s), q(t_0 - s)\big) \quad \text{for all } s \in \mathbb{R}.$$

11.1.3 Reachability

By an appropriate choice of the initial condition, one can get from any position in space to any other.[3]

11.6 Theorem (Reachability)
If the connected Riemannian manifold (M, g) *is geodesically complete, then the Hamiltonian dynamics generated by* (11.1.5) *satisfies this property: For all energies* $E > \sup_q V(q)$ *and positions* $q_0, q_1 \in M$*, there is an initial condition* $x_0 = (p_0, q_0) \in \Sigma_E$ *and a time* $t \geq 0$ *such that* $q(t, x_0) = q_1$*.*

Proof: According to Theorem 8.31, the solution curves of the Hamiltonian equations coincide, up to parametrization, with the geodesics of the Jacobi metric

$$g_E(q) = \big(E - V(q)\big)\, g(q) \qquad \big(q \in M\big),$$

analogous to Definition 8.30. So one shows that a geodesic segment of (M, g_E) connecting q_0 and q_1 exists. This however is guaranteed by the Hopf-Rinow theorem (Theorem G.15, Criterion 2. on page 589), because the Riemannian manifold $\big(M, g_E\big)$ is geodesically complete due to the bound $g_E \geq \big(E - \sup_q V(q)\big) g$. □
As an example, the statement holds true for Hamiltonians of the form (11.1.1). This property also distinguishes motion in a potential from motion in a magnetic field: Planar motion in a constant magnetic field is circular (see Section 6.3.3), so it cannot connect points with too large a distance.

11.2 Motion in a Periodic Potential

In the following, we consider the motion of a *single* particle on $\mathbb{R}^d$. So there is only a single mass m occurring in the kinetic energy term. Multiplying the Hamiltonian by m only rescales the time. So we can simply study the Hamiltonian differential equation

$$\dot{p} = -\nabla V(q) \quad , \quad \dot{q} = p \, . \tag{11.2.1}$$

generated by

$$H : P \to \mathbb{R} \quad , \quad H(p, q) = \tfrac{1}{2}\|p\|^2 + V(q) \tag{11.2.2}$$

on the phase space $P := \mathbb{R}^d_p \times \mathbb{R}^d_q$. We will assume that the potential $V \in C^2\big(\mathbb{R}^d_q, \mathbb{R}\big)$ is $\mathcal{L}$*-periodic* with respect to a *lattice*

$$\mathcal{L} := \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_d) = \Big\{ \textstyle\sum_{i=1}^d n_i \ell_i \;\Big|\; n_i \in \mathbb{Z} \Big\}$$

[3] In this, it is not essential for the flow to be complete, i.e., whether the potential V satisfies condition (11.1.3). It is however essential for the energy E to be sufficiently large, compare Exercise 6.34.2.

spanned by the basis vectors $\ell_1, \ldots, \ell_d$ of $\mathbb{R}^d$; this means that

$$V(q+\ell) = V(q) \qquad \big(q \in \mathbb{R}^d,\ \ell \in \mathcal{L}\big).$$

This differential equation models for example the motion of a classical electron in a d-dimensional crystal.

11.2.1 Existence of Asymptotic Velocities

As an application of Birkhoff's ergodic Theorem 9.32, we compare the asymptotics of the motion in future and past. λ^{2d} will denote the Lebesgue measure on the $2d$-dimensional phase space $P = \mathbb{R}^d_p \times \mathbb{R}^d_q$.

11.7 Theorem (Asymptotic Velocities in a Periodic Potential)

- *The ODE (11.2.1) generates a flow* $\Phi \equiv (p, q) \in C^1(\mathbb{R} \times P, P)$.
- *For* λ^{2d}*-almost all initial conditions* $x \in P$*, the* **asymptotic velocities**

$$\overline{v}^{\pm}(x) := \lim_{T \to +\infty} \frac{1}{T} \int_0^T p(\pm t, x)\ \mathrm{d}t$$

exist. Setting $\overline{v}^{\pm}(x) := 0$ *otherwise, one obtains measurable mappings*

$$\overline{v}^{\pm} : P \to \mathbb{R}^d .$$

For λ^{2d}*-almost all* x*, one has* $\overline{v}^{+}(x) = \overline{v}^{-}(x)$.
- *One has* $\|\overline{v}^{\pm}(x)\| \le \sqrt{2(H(x) - V_{\min})}$*, with* $V_{\min} := \inf_{q \in \mathbb{R}^d} V(q) > -\infty$.

11.8 Remark
The conclusion that the asymptotic velocities of past and future coincide almost always is actually quite surprising, because for chaotic dynamical systems, the far future cannot be predicted from a finite precision knowledge of the past. $\Diamond$

Birkhoff's ergodic theorem, which is used in the existence proof of $\overline{v}^{\pm}$, requires a finite measure, but λ^{2d} fails to satisfy this hypothesis. For this reason, we prepare by first constructing a comparison dynamics on a compact space.

Due to the $\mathcal{L}$-periodicity, we can also view V as a function on the d-dimensional torus

$$\mathbb{T} := \mathbb{R}^d/\mathcal{L} = \left\{ q + \mathcal{L} \mid q \in \mathbb{R}^d \right\}.$$

This manifold can be identified with the compact parallelotope

$$D := \left\{ \textstyle\sum_{i=1}^d x_i \ell_i \;\middle|\; x_i \in [0,1] \right\} \subset \mathbb{R}^d ,$$

called the *elementary domain*, after identifying opposite parts of the boundary by means of the equivalence relation $q \sim r$ when $q - r \in \mathcal{L}$. In particular, $\mathbb{T}$ is compact. Intuitively speaking, the smooth mapping

$$\pi : \mathbb{R}^d_q \to \mathbb{T} \quad , \quad q \mapsto q + \mathcal{L}$$

wraps the configuration space onto the torus and allows us to define the potential

$$\widehat{V} : \mathbb{T} \to \mathbb{R} \quad , \quad \widehat{V} = V \circ \pi^{-1}$$

on $\mathbb{T}$. The phase space of the torus is the $2d$-dimensional manifold

$$\widehat{P} := \mathbb{R}^d_p \times \mathbb{T} .$$

The mapping $\widehat{\pi} : P \to \widehat{P}$, $(p, q) \mapsto \big(p, \pi(q)\big)$ of the phase spaces is a local diffeomorphism. The projection of the function (11.2.2) onto $\widehat{P}$, i.e.,

$$\widehat{H} : \widehat{P} \to \mathbb{R} \quad , \quad \widehat{H}(p, q) = \tfrac{1}{2}\|p\|^2 + \widehat{V}(q) , \tag{11.2.3}$$

has as its Hamiltonian differential equation $\dot{p} = -\nabla V(q)$, $\dot{q} = p$; its solution is given by a Hamiltonian flow $\widehat{\Phi} = (\widehat{p}, \widehat{q}\,) : \mathbb{R} \times \widehat{P} \to \widehat{P}$.

Proof of Theorem 11.7:

• First note that $V_{\min} > -\infty$ and $V_{\max} := \sup_{q \in \mathbb{R}^d} V(q) < +\infty$, because $\inf_{q \in \mathbb{R}^d} V(q) = \inf_{q \in D} V(q)$, and D is compact and V is continuous.

Thus the hypothesis of Theorem 11.1 is satisfied, and the differential equation (11.2.1) generates a complete flow $\Phi \in C^1(\mathbb{R} \times P, P)$.

• Because $\hat{\pi} \circ \Phi_t = \widehat{\Phi}_t \circ \hat{\pi}$, it follows (with the stipulation $\overline{v}^{\pm}(\hat{x}) := 0$ in case the limit fails to exist) that

$$\overline{v}^{\pm}(x) = \overline{v}^{\pm}(\hat{x}) := \lim_{T \to +\infty} \frac{1}{T} \int_0^T \widehat{p}\,(\pm t, \hat{x})\, \mathrm{d}t \quad \text{for} \quad \hat{x} := \hat{\pi}(x) . \tag{11.2.4}$$

• With the measure $\hat{\lambda} := \lambda^d \times \mu$ on $\widehat{P} = \mathbb{R}^d_p \times \mathbb{T}$ and the Haar measure μ on the torus normalized by $\mu(\mathbb{T}) = \lambda^d(D)$, $\hat{\lambda}$ is invariant under $\widehat{\Phi}$, and for all measurable subsets $A \subseteq P$, one has

$$\lambda^{2d}\big(A\big) = \sum_{\ell \in \mathcal{L}} \lambda^{2d}\big(A \,\cap\, \mathbb{R}^d_p \times (D + \ell)\big) = \sum_{\ell \in \mathcal{L}} \hat{\lambda}\Big(\hat{\pi}\big(A \,\cap\, \mathbb{R}^d_p \times (D + \ell)\big)\Big) , \tag{11.2.5}$$

because the translates $D + \ell \quad (\ell \in \mathcal{L})$ of the elementary domain have $\mathbb{R}^d$ as their union, and their intersection is of measure 0. So it suffices to show that for $\hat{\lambda}$-almost all $\hat{x} \in \widehat{P}$, the limits $\overline{v}^{\pm}(\hat{x})$ exist, and $\overline{v}^{+}(\hat{x}) = \overline{v}^{-}(\hat{x})$.

• Now $\hat{\lambda}$ still is not a finite measure, but for all $E \in \mathbb{R}$, the restriction of $\hat{\lambda}$ to the $\widehat{\Phi}$-invariant sublevel set

$$\widehat{P}_E := \left\{ \hat{x} \in \widehat{P} \mid \widehat{H}(\hat{x}) \le E \right\}$$

of the phase space is a finite measure, with $\hat{\lambda}(\widehat{P}_E) \le \lambda^d(B_r^d)\,\mu(\mathbb{T}) < \infty$ for the d-dimensional ball B_r^d of radius $r := \sqrt{2(E - V_{\min})}$.
On the other hand, $\hat{\lambda}(\widehat{P}_E) > 0$ for $E > V_{\max}$, with $V_{\max} = \sup_q V(q) < \infty$. We can therefore normalize $\hat{\lambda}\restriction_{\widehat{P}_E}$ to a probability measure and apply Birkhoff's ergodic Theorem 9.32 to it. Null sets remain null sets if we merely multiply the measure by a positive constant. So we conclude

$$\overline{v}^{+}(\hat{x}) = \overline{v}^{-}(\hat{x}) \quad \text{for} \quad \hat{\lambda} - \text{ almost all } \hat{x} \in \widehat{P}_E\,.$$

Since $\widehat{P} = \bigcup_{E \in \mathbb{N}} \widehat{P}_E$, the same conclusion now follows even for $\hat{\lambda}$-almost all $\hat{x} \in \widehat{P}$. The set of pre-images of $\hat{x} \in \widehat{P}$ under $\hat{\pi} : P \to \widehat{P}$ is countable, like the lattice $\mathcal{L}$. Because of (11.2.4) and (11.2.5), it also follows that $\overline{v}^{+}(x) = \overline{v}^{-}(x)$ for λ^{2d}-almost all $x \in P$, and the measurability of the mappings $\overline{v}^{\pm}$.
• As $\|p\| \le \sqrt{2(E - V_{\min})}$ for all $(p, q) \in \Sigma_E = H^{-1}(E)$, the Cesàro-means (11.2.3) have the same majorant. □

11.9 Remark (Random Potentials)
In the theory of metallic alloys, one assumes that the sites of a lattice $\mathcal{L} \subset \mathbb{R}^d$ are occupied randomly by atoms from one out of two or more chemical elements. In this case one refers to this as a *substitutional alloy*. Assigning to an atom of kind $i \in \mathcal{I}$ a compactly supported *single site potential* $W_i \in C_c^2(\mathbb{R}^d, \mathbb{R})$, and making a choice $\omega \in \Omega := \mathcal{I}^{\mathcal{L}}$ of the kind of atom at each site, one obtains the total potential as

$$V : \Omega \times \mathbb{R}^d \to \mathbb{R} \quad , \quad V(\omega, q) := \textstyle\sum_{\ell \in \mathcal{L}} W_{\omega(\ell)}(q - \ell)\,.$$

To describe a particular alloy, one chooses a probability measure β on Ω that is invariant under translations from $\mathcal{L}$. In the simplest case, this is the product measure of the probability measure on $\mathcal{I}$ that quantifies the mixing ratio of the alloy.

The Hamiltonian $H : \Omega \times T^*\mathbb{R}^d \to \mathbb{R}$, $H(\omega; p, q) = \frac{1}{2}\|p\|^2 + V(\omega, q)$ on the extended phase space defines a time evolution on $T^*\mathbb{R}^d$ parametrized by Ω. The asymptotic velocity exists for $\beta \times \lambda^{2d}$-almost all initial conditions on $\Omega \times T^*\mathbb{R}^d$, and their *distribution* is the same for β-almost all Hamiltonians $H(\omega; \cdot)$ $(\omega \in \Omega)$.[4] The motion in such a random potential is shown at the beginning of the chapter, on page 33, underlaid by a gray scale image of the realization $V(\omega, \cdot)$. ♢

[4]Provided β is ergodic under translations from $\mathcal{L}$, as is indeed the case e.g. for the product measure.

11.2.2 Distribution of Asymptotic Velocities

Which asymptotic velocities do typically occur? As the limits $\overline{v}^+$ and $\overline{v}^-$ anyways coincide almost everywhere, we simply consider the measurable mapping

$$\overline{v} : \widehat{P} \to \mathbb{R}^d \ , \ \overline{v}(\hat{x}) := \begin{cases} \overline{v}^+(\hat{x}) & \text{if } \overline{v}^+(\hat{x}) = \overline{v}^-(\hat{x}) \\ 0 & \text{otherwise} . \end{cases}$$

Referring to the Hamiltonian $\widehat{H}$ from (11.2.3), we obtain the *energy momentum mapping*

$$I := \left(\widehat{H}, \overline{v}\right) : \widehat{P} \longrightarrow \mathbb{R} \times \mathbb{R}^d .$$

The image measure $\nu := I(\hat{\lambda})$ on $\mathbb{R} \times \mathbb{R}^d$ describes the joint distribution of energy and asymptotic velocity.

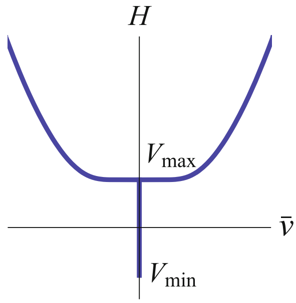
11.10 Examples (Energy Momentum Mapping)

1. For the potential $V = 0$ and $\hat{x} = (\hat{p}, \hat{q})$, one has $\overline{v}(\hat{x}) = \hat{p}$ and $\widehat{H}(\hat{x}) = \frac{1}{2}\|\hat{p}\|^2$.

   Therefore the support of ν is the paraboloid

   $$\text{supp}(\nu) = \left\{(E, \overline{v}) \in \mathbb{R}\times\mathbb{R}^d \mid E = \tfrac{1}{2}\|\overline{v}\|^2\right\} .$$

   For a fixed energy $E > 0$, the asymptotic velocity is therefore uniformly distributed on a sphere of radius $\sqrt{2E}$ in $\mathbb{R}^d$.
2. For $d = 1$, the support of ν bifurcates at energy $V_{\max}$, see Theorem 11.11. The figure on the right shows the support of ν for the potential $V = \cos$. ◊



In space dimension 1, the asymptotic velocity exists always and can actually be calculated:

11.11 Theorem *If $d = 1$ and ℓ is a period of the lattice, one has the estimates:*

1. *For $E \leq V_{\max}$ and initial condition $x_0 = (p_0, q_0) \in H^{-1}(E)$:*

   $$\left|q(t; x_0) - q_0\right| \leq \ell \qquad (t \in \mathbb{R}).$$

2. *For $E > V_{\max}$ and initial condition $x_0 = (p_0, q_0) \in H^{-1}(E)$:*

   $$\left|q(t; x_0) - (q_0 + \overline{v}(x_0)t)\right| \leq \ell \qquad (t \in \mathbb{R}),$$

   with the **asymptotic velocity**

$$\overline{v}(x_0) := \lim_{T\to\infty} \frac{1}{T}\int_0^T p(t;x_0)\,\mathrm{d}t = \frac{\mathrm{sign}(p_0)}{\ell^{-1}\int_0^\ell \bigl(2(E-V(q)\bigr)^{-\frac{1}{2}}\,\mathrm{d}q}\,.$$

Proof:

- To begin with, it is useful to find the phase portrait, i.e., the decomposition of the phase space into orbits.
  As the orbit through $x_0 \in \Sigma_E := H^{-1}(E)$ is contained in Σ_E, for regular values E of H, every connected component of this level line of H will be an orbit. As the term $\frac{1}{2}p^2$ in the sum $H(p,q)$ has only one value that is not regular, namely 0, the regular values of H and V coincide. In particular, all values $E > V_{\max}$ are regular.
- For $E < V_{\max}$, the particle cannot be at positions q with $V(q) > E$. These prohibited zones are nonempty intervals that are arranged with the periodicity of the lattice. So their complement $\{q \in \mathbb{R} \mid V(q) \le E\}$ is the disjoint union of closed intervals of length $< \ell$. This proves the first part of the theorem for $E < V_{\max}$.
  If $E = V_{\max}$, then in both directions within distance $\le \ell$ from q_0, we find points q_1, q_2 with $V(q_1) = V(q_2) = V_{\max}$. The points $(0,q_1), (0,q_2) \in \Sigma_E$ are equilibria, and the orbit through x_0 lies between them. This means that the motion is bound for $E = V_{\max}$ as well.

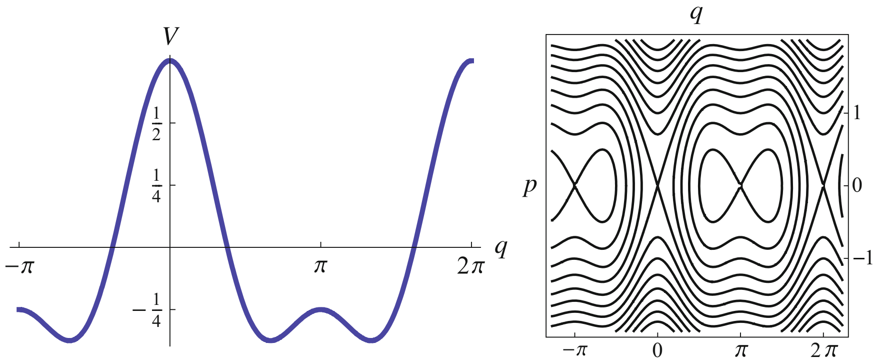


Figure 11.2.1 The potential $V(q) = \frac{1}{2}\cos(q) + \frac{1}{4}\cos(2q)$ (left) and the corresponding phase portrait of $H(p,q) = \frac{1}{2}p^2 + V(q)$ (right)

- In the second case, $E > V_{\max}$, the set Σ_E consists of only two regular connected components, namely orbits, which can be written as the graphs of

$$p_E^{\pm} : \mathbb{R} \to \mathbb{R} \quad , \quad p_E^{\pm}(q) := \pm\sqrt{2(E - V(q))}\,.$$

  Hereby, $p_E^+ > 0$ corresponds to a motion to the right, and $p_E^- < 0$ to a motion to the left. Then

$$T := \int_0^\ell \left|\frac{\mathrm{d}t}{\mathrm{d}q}\right| \mathrm{d}q = \int_0^\ell \frac{1}{p_E^+(q)}\,\mathrm{d}q = \int_0^\ell \frac{1}{\sqrt{2(E-V(q))}}\,\mathrm{d}q$$

is the time needed by the particle of energy E to proceed by one lattice period ℓ. This is where the ansatz for the average velocity $\overline{v} = \frac{\ell}{T}$ comes from. Within time $t > 0$, the particle proceeds by at least $n := \lfloor t/T \rfloor \in \mathbb{N}_0$, and by at most $n+1$ lattice periods. This proves the claim (Figure 11.2.1). □

One dimensional motion is not particularly interesting from the point of view of physics. We can reduce the motion in a periodic potential in $\mathbb{R}^d$ to the motion in 1-dimensional potentials, if after choosing an appropriate orthogonal basis, V can be written in the form

$$V : \mathbb{R}^d \to \mathbb{R} \quad , \quad V(q) = \sum_{i=1}^d V_i(q_i)\,, \tag{11.2.6}$$

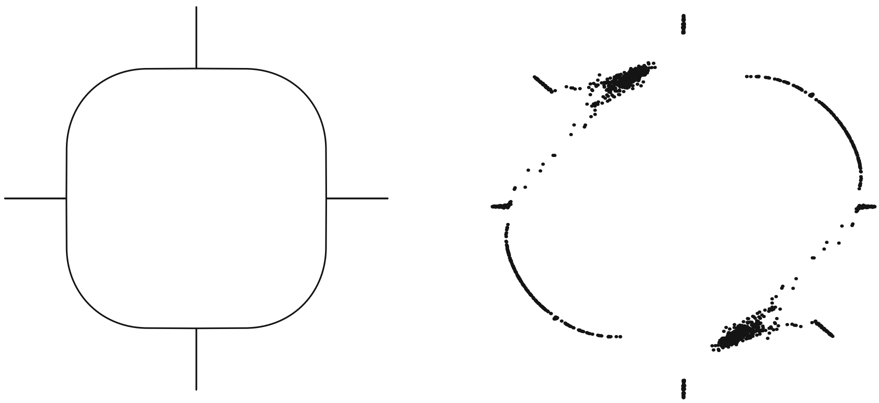


Figure 11.2.2 Distribution of asymptotic velocities for energy $E = 3$ and potentials $V(q_1, q_2) = \cos(q_1) + \cos(q_2)$ (left) and $V(q_1, q_2) = \cos(q_1) + \cos(q_1 + q_2)$ (right)

see Figure 11.2.2, left. Such potentials are called *separable*. In general, this is however not the case, for instance not for the potential $V(q_1, q_2) = \cos(q_1) + \cos(q_1 + q_2)$. Its distribution of asymptotic velocities can be seen in Figure 11.2.2, right. One observes in these pictures that the resulting asymptotic velocities, which make up the support of the measure ν, are symmetric with respect to the reflection

$$S : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R} \times \mathbb{R}^d \quad , \quad (E, \overline{v}) \mapsto (E, -\overline{v})\,.$$

This is true for arbitrary potentials, and more generally:

11.12 Lemma *The distribution ν of energy and asymptotic velocity is S-invariant.*

Proof:

- If the limit $\overline{v}^+(x)$ of the asymptotic velocity of $x \in P$ exists, then from the reversibility of the flow Φ by Theorem 11.5, it follows for the time-reversed initial condition $\mathcal{T}(x)$ that

$$\overline{v}^-\big(\mathcal{T}(x)\big) = \lim_{T\to+\infty} \frac{1}{T}\int_0^T p\big(-t, \mathcal{T}(x)\big)\,\mathrm{d}t = \lim_{T\to+\infty} \frac{-1}{T}\int_0^T p(+t, x)\,\mathrm{d}t = -\overline{v}^+(x).$$

  Otherwise, by definition of $\overline{v}^\pm$, both sides equal zero.
  By the semiconjugacy property $\hat{\pi} \circ \Phi_t = \hat{\Phi}_t \circ \hat{\pi}$ of the flows, this relation also holds on the phase space $\widehat{P}$, because $\overline{v}^-\big(\pi(\mathcal{T}(x))\big) = -\overline{v}^+\big(\pi(x)\big)$.
- By Theorem 11.7, $\overline{v}^+$ and $\overline{v}^-$ coincide $\hat{\lambda}$-almost everywhere with $\overline{v}$. For the energy-momentum mapping $I = \big(\widehat{H}, \overline{v}\big)$ and time reversal $\widehat{\mathcal{T}}$ on $\widehat{P}$, one has therefore

$$S \circ I(\hat{x}) = I \circ \widehat{\mathcal{T}}(\hat{x}) \qquad (\hat{\lambda} - \text{almost everywhere}). \tag{11.2.7}$$

- The measure $\hat{\lambda}$ is $\widehat{\mathcal{T}}$-invariant. Since ν was defined as the image measure of $\hat{\lambda}$ under the energy-momentum mapping, this, in connection with (11.2.7), proves the claim. □

11.2.3 Ballistic and Diffusive Motion

11.13 Definition *The point*[5] $x_0 \in P$ *in phase space is called* **ballistic**, *if its asymptotic velocity* $\overline{v}(x_0)$ *is non-zero.*

If, in contrast, $\overline{v}(x_0)$ equals 0, this does not automatically mean that the motion is bound, i.e., that $q(t, x_0)$ stays for all times t in a bounded domain of the configuration space $\mathbb{R}^d$. As we will see, there are also cases in which $\|q(t, x_0) - q_0\|$ typically diverges like $\sqrt{|t|}$ rather than $|t|$. Such motions are called *diffusive*.

As a simple standard of reference for the dynamics, we use free motion, which is generated by the purely kinetic Hamiltonian $H(p, q) = \frac{1}{2}\|p\|^2$,

$$\big(p(t, x_0),\ q(t, x_0)\big) = (p_0,\ q_0 + p_0\, t) \qquad \big(x_0 = (p_0, q_0) \in P,\ t \in \mathbb{R}\big).$$

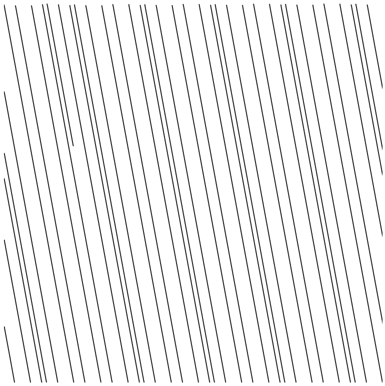
For this motion, we can choose an arbitrary lattice $\mathcal{L} \subset \mathbb{R}^d_q$. The following results hold in this case:

[5] Then the orbit $\mathcal{O}(x_0)$ will also be called *ballistic*, because the asymptotic velocity is clearly the same for all points of $\mathcal{O}(x_0)$.

1. For positive energies, the motion is *ballistic*. For all $x_0 \equiv (p_0, q_0) \in P$, one has $\overline{v}(x_0) = p_0$, hence for $p_0 \neq 0$,

$$\lim_{|t|\to\infty} \frac{\|q(t, x_0) - q_0\|}{|t|} = \|p_0\| > 0\,.$$

2. *Integrability:* For $E > 0$, the energy surface $\widehat{H}^{-1}(E) \subset \widehat{P}$ is fibered by flow-invariant d-dimensional tori. In this case, the phase space tori are of the form $\mathbb{T}_p := \{p\} \times \mathbb{T}$ with $\|p\| = \sqrt{2E}$ and with $\mathbb{T} = \mathbb{R}^d/\mathcal{L}$ being the torus in configuration space.
So the system is completely integrable.



We now study which of these properties stay true for motion in a periodic potential.

11.14 Examples ($d = 1$) In the case of a periodic potential $V \in C^2(\mathbb{R}, \mathbb{R})$ (Theorem 11.11), both characteristics of free motion remain true for all energies $E > V_{\max}$: The limit $\overline{v}(x_0)$ exists and is non-zero, and the energy shell $\widehat{H}^{-1}(E)$ is the disjoint union of (exactly two) invariant one-dimensional tori. ♢

In higher dimensions, the ballistic character of the motion is subtle:

11.15 Exercise (Ballistic and Bound Motion) Show:

(a) Let the dimension be $d \in \mathbb{N}$, and $E > V_{\max}$. Then for every lattice vector $\ell \in \mathcal{L} \setminus \{0\}$, there exist initial conditions $x_0 \in \Sigma_E$ with asymptotic velocity

$$\overline{v}(x_0) \neq 0 \quad \text{and direction} \quad \frac{\overline{v}(x_0)}{\|\overline{v}(x_0)\|} = \frac{\ell}{\|\ell\|}\,.$$

So initial conditions whose solution curves are ballistic *do exist*.

(b) In space dimension $d > 1$, there exist periodic potentials that lead to *bound* orbits, even for appropriate initial conditions $x_0 \in \Sigma_E$ with certain energies $E > V_{\max}$:

$$\sup_{t\in\mathbb{R}} \|q(t, x_0)\| < \infty\,.$$

So the distance from the start point doesn't scale like t^1, but like t^0. ♢

More realistic periodic potentials in physics contain coulombic singularities at the sites s of the nuclei. There, the potential $V(q)$ is asymptotic to $\frac{-z}{\|q-s\|}$, where $z > 0$ denotes the charge of the nucleus. As an example, we consider the periodic potential

$$V(q) := \sum_{s\in\mathbb{Z}^2} \frac{-e^{-\|q-s\|}}{\|q-s\|} \qquad \left(q \in \mathbb{R}^2 \setminus \mathbb{Z}^2\right)$$

in the plane, which is obtained by adding up local Yukawa potentials at all sites $s \in \mathbb{Z}^2$ of nuclei in the plane. For the Hamiltonian H with this potential, the following theorem holds:

11.16 Theorem (Deterministic Diffusion **[Kn1]**)

There exists a threshold energy $E_0 > 0$ such that for all $E > E_0$ and all those probability measures μ_E with support in the energy shell Σ_E that are absolutely continuous with respect to the Liouville measure (see page 193) and that have sufficient spatial decay to ensure $\langle \|q_0\|^2 \rangle_{\mu_E} < \infty$, the following limit exists:

$$D(E) := \lim_{t \to \infty} \frac{1}{t} \left\langle \|q(t) - q_0\|^2 \right\rangle_{\mu_E} > 0 \,.$$

Here the bracket $\langle \, \rangle_{\mu_E}$ denotes the expected value with respect to μ_E.

As a matter of fact, the constant D depends only on E, but not on the choice of measure μ_E. D can be interpreted as a diffusion constant.

We take the expected value in order not to have to take exceptional orbits, like ballistic or bound orbits, into explicit consideration. (Both of these types do occur, but have measure 0.)

In the long term limit, typical orbits behave like paths of Brownian motion, see Figure 11.2.3.

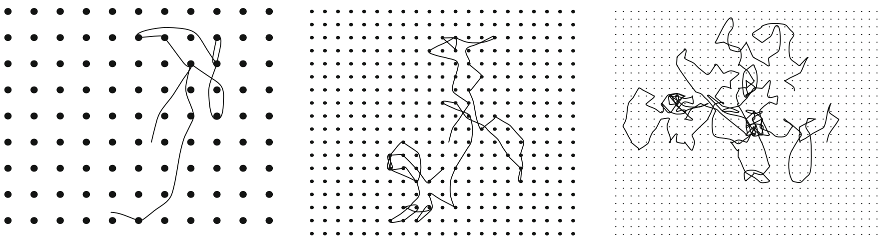


Figure 11.2.3 Orbit in a periodic coulombic potential for total time $T = 1$ (left), $T = 4$ (center) and $T = 16$ (right). The length has been scaled like $1/\sqrt{T}$

11.17 Remark (Anomalous Diffusion)
We denote as *anomalous diffusion* the case when for some $\alpha \in (0, 2)$, the limit

$$\lim_{t \to \infty} \frac{1}{t^\alpha} \|q(t, x_0) - q_0\|^2$$

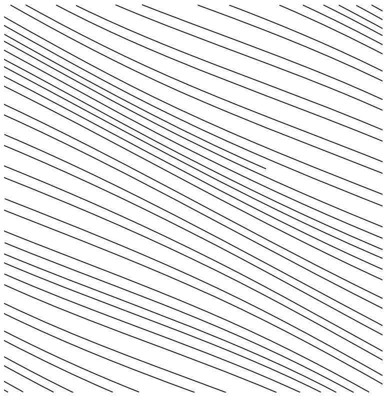
exists and is positive. Namely *subdiffusion* refers to the case $\alpha \in (0, 1)$ and *superdiffusion* to $\alpha \in (1, 2)$. For ordinary diffusion one has $\alpha = 1$. In Geisel, Zacherl and Radons [GZR], the occurrence of anomalous diffusion is shown numerically for periodic smooth potentials, and the value of α is explained by the coexistence of KAM tori and ergodic domains in phase space. ◊

Finally, we want to remark that for *smooth* periodic potentials, classical motion at *high* energies is typically ballistic.

More precisely, there exists a subset of the energy shell $\widehat{H}^{-1}(E)$ (with the Hamiltonian $\widehat{H}$ from 11.2.3), whose (Liouville) measure for high energies E is asymptotic to the measure of the energy shell, such that the motion with initial conditions in this set is ballistic. The reason is that for high energies, the smooth periodic potential can be viewed as a small perturbation of the case $V = 0$, namely of free motion.

The KAM theorem (Kolmogorov, Arnol'd and Moser), see Theorem 15.33, guarantees in this case that many of the invariant phase space tori of free motion will only be deformed but not destroyed (here we consider motion within a given energy shell $\widehat{H}^{-1}(E) \subset \widehat{P}$). This will be shown in Example 15.33.

But if we have this kind of motion (on KAM tori) over a torus $\mathbb{T} = \mathbb{R}^d/\mathcal{L}$ in configuration space (see the figure), then the motion in $\mathbb{R}^d$ is ballistic.



This in turn leads to the question whether the energy shell $\widehat{H}^{-1}(E)$ for energies $E > V_{\max}$ consists entirely of invariant tori, in other words whether the motion is *completely integrable*.

We know that this is the case for one degree of freedom (Example 11.14). In contrast, we have the following result [Kn2]:

11.18 Theorem (Complete Integrability)
Let the potential $V \in C^2(\mathbb{R}^d, \mathbb{R})$ be $\mathcal{L}$-periodic and $d \geq 2$. If there exists an energy surface $\widehat{H}^{-1}(E)$ with $E > V_{\max}$ on which the motion is completely integrable, then the potential is constant.

11.19 Exercise (Complete Integrability)
Show Theorem 11.18 for separable potentials, i.e., those of the form (11.2.6). ◊

11.20 Remark (Quantum Mechanics) In the case of a Schrödinger operator

$$-\tfrac{\hbar^2}{2}\Delta + V \qquad \text{on the Hilbert space } L^2(\mathbb{R}^d)\,,$$

the motion in a periodic potential V is *always* ballistic, even for small energies. In this situation, the mathematical question arises why and in what manner the *correspondence principle* is violated. This heuristic law states that quantum mechanical motion should become 'similar' to classical motion when the Planck constant $\hbar$ is small. What is happening here is the exchange of two limits, namely the semiclassical limit $\hbar \searrow 0$ and the time limit $t \to \infty$; such exchanges are not generally permissible.

In physics, the question about the type of motion (bound, ballistic, diffusive etc.) is interesting when considering transport phenomena in a solid. ◊

11.3 Celestial Mechanics

"Mathematical physics, as we are well aware, is an offspring of celestial mechanics." Henri Poincaré, in [Poi3]

The n-body problem in celestial mechanics has a special status among Hamiltonian dynamics in a potential: it is the longest and most intensely studied among these differential equations.

The equations of motion (1.8) of the n-body problem in d space dimensions are (the version written in 2nd order of) the Hamiltonian differential equations for the Hamiltonian

$$H : \widehat{P} \to \mathbb{R}\,,\, H(p,q) = \sum_{k=1}^{n} \frac{\|p_k\|^2}{2m_k} + V(q) \quad \text{with} \quad V(q) := -\sum_{1\le k<\ell\le n} \frac{m_k m_\ell}{\|q_k - q_\ell\|} \tag{11.3.1}$$

on the phase space $\widehat{P} := T^*\widehat{M} \cong \mathbb{R}^{dn} \times \widehat{M}$ over the configuration space

$$\widehat{M} := \left\{ q = (q_1, \ldots, q_n) \in \mathbb{R}^{dn} \mid q_k \neq q_\ell \text{ for } k \neq \ell \right\}. \tag{11.3.2}$$

The masses $m_k > 0$ are parameters in the differential equation. The *total mass* is denoted by $m_N := \sum_{k=1}^{n} m_k$.

11.21 Exercise (Constants of Motion)
Show that in d space dimensions, along with H, the following phase space functions are constants of the motion generated by (11.3.1) in the phase space $\widehat{P}$ (if need be extended with the time axis $\mathbb{R}_t$):

- The **total momentum** $p_N : \widehat{P} \to \mathbb{R}^d \quad , \quad (p,q) \mapsto \sum_{k=1}^{n} p_k$,
- **Center of Mass at time 0** (see also Theorem 12.38):

$$q_N : \widehat{P} \times \mathbb{R}_t \to \mathbb{R}^d \quad , \quad (p,q;t) \mapsto \tfrac{1}{m_N} \sum_{k=1}^{n} (m_k q_k - p_k t)\,,$$

- The **total angular momentum** with the $\binom{d}{2}$ components

$$L_{i,j} : \widehat{P} \to \mathbb{R}\,,\, (p,q) \mapsto \sum_{k=1}^{n} \left(q_{k,i}\, p_{k,j} - q_{k,j}\, p_{k,i} \right) \qquad (1 \le i < j \le d).$$

How many of these altogether $\binom{d+2}{2}$ constants of motion are algebraically independent in $d = 2$ or $d = 3$ dimensions respectively? ◊

11.3.1 Geometry of the Kepler Problem

In the introduction, we proved that the orbits of the Kepler problem in configuration space are conics. So for negative energy E, the mass point moves on elliptical orbits around the origin, whereas for positive E, it is led on a hyperbola. The case $E = 0$ corresponds to parabolas.

So initially it appears that we have completely understood the Kepler problem (and thus also the dynamics of two bodies), at least up to the time parametrization. And once the shape of the orbit is known, the time parametrization can be found by integration.

At closer look, however, the Kepler dynamics offers some surprises and special features. One needs to know these in order to better understand the n-body problem, or the perturbation theory of the Kepler problem.

Let us begin with the Hamiltonian version: On the configuration space[6] $\widehat{M} := \mathbb{R}^d \backslash \{0\}$, the Kepler, or Coulomb potential $V : \widehat{M} \to \mathbb{R}$, $V(q) = \frac{-Z}{\|q\|}$ is defined. In the case of gravitational (attractive) forces, $Z > 0$ is the (reduced) mass of the central force. In the electrostatic case, $-Z$ is the product of both charges. So in the case of charges of equal sign, one has repulsive forces.

In any case, the phase space is of the form

$$\widehat{P} := T^*\widehat{M} \cong \mathbb{R}^d \times \widehat{M},$$

and as $\widehat{P} \subset T^*\mathbb{R}^d$ is open, the restriction $\hat{\omega} := \omega_0 \restriction_{\widehat{P}}$ of the canonical symplectic form $\omega_0 := \sum_{i=1}^d dq_i \wedge dp_i$ on $T^*\mathbb{R}^d$ is symplectic (see Definition 10.3 on page 217). The motion generated by the Hamiltonian

$$\hat{H} : \widehat{P} \to \mathbb{R} \quad , \quad \hat{H}(p,q) = \tfrac{1}{2}\|p\|^2 + V(q) \tag{11.3.3}$$

is complete for $Z < 0$, because only energies $E > 0$ occur, and for these, the minimal distance $\min\{\|q\| \mid \exists p \in \mathbb{R}^d : \hat{H}(p,q) = E\}$ is $-Z/E > 0$.

For the case $Z > 0$, which we will now study further, the maximal flow

$$\hat{\Phi} : D \to \widehat{P} \quad \text{with domain} \quad D \subset \mathbb{R}_t \times \widehat{P} \tag{11.3.4}$$

(see Theorem 3.39) is incomplete for exactly those initial values $x_0 = (p_0, q_0) \in \widehat{P}$ for which the initial momentum p_0 is contained in $\mathrm{span}(q_0)$. We will regularize the corresponding collision orbits in a moment.

As the trajectory $t \mapsto q(t, x_0) \in \widehat{M}$ remains in the subspace spanned by q_0 and p_0, we may assume without loss of generality that the dimension d of the configuration space $\widehat{M}$ equals 2, and we set $\ell := \hat{L}(x_0)$ with angular momentum

$$\hat{L} : \widehat{P} \to \mathbb{R} \quad , \quad \hat{L}(p,q) = q_1 p_2 - q_2 p_1 . \tag{11.3.5}$$

As already shown in (1.4), $\hat{L}$ is constant in time.

The simplest shape of an orbit is a *circle*, which occurs (as a special case of an ellipse) only for energies $E := H(x_0) < 0$, namely when $E = \frac{-Z}{2\|q_0\|}$. This can be

[6] While the configuration space in physics is 3-dimensional, the case $d = 2$ is also relevant because motion in a central force field is planar.

seen by noting that the effective potential $V_\ell(r) = \frac{\ell^2}{2r^2} - \frac{Z}{r}$ has its minimum for $r := \|q_0\|$, where $\ell^2 = rZ$, and the value of the minimum is E. For all other initial conditions (p_0, q_0) and corresponding constants of motion (E, ℓ), the distance of the *pericenter*, which is the orbit point closest to the center, $r_{\min} : \widehat{P} \to [0, \infty)$, is

$$r_{\min}(p_0, q_0) = \inf\left\{r > 0 \mid V_\ell(r) \le E\right\} = \begin{cases} \frac{-Z+\sqrt{Z^2+2E\ell^2}}{2E} & , E \neq 0 \\ \frac{\ell^2}{2Z} & , E = 0 \end{cases} ,$$

which is not a minimum of V_ℓ. This minimum distance is 0 if and only if the angular momentum vanishes, too. Each orbit takes on its pericenter distance at least once (where this distance corresponds to a collision in case $\ell = 0$). If the orbit is not a circle, it can be parametrized locally by the time $\hat{T} : \widehat{P} \to \mathbb{R}$ that has to elapse until the next pericenter. As the radial velocity is of magnitude $\sqrt{2E + 2\frac{Z}{R} - \frac{\ell^2}{R^2}}$ (see (1.6)), it follows that

$$\hat{T}(p, q) = \int_{r_{\min}}^{r} \frac{1}{\mathrm{d}R/\mathrm{d}t}\,\mathrm{d}R = \operatorname{sign}(\langle p, q\rangle) \int_{r_{\min}}^{r} \frac{R}{\sqrt{2R^2E + 2ZR - \ell^2}}\,\mathrm{d}R\,. \quad (11.3.6)$$

For positive E, we evaluate this integral in Exercise 12.10.

As for the properties discussed so far, the Kepler problem hardly differs from other homogenous singular potentials $q \mapsto -Z/\|q\|^a$ with $a > 0$. The special nature of the exponent $a = 1$ lies in the existence of an additional conserved quantity:

11.22 Exercise (Laplace-Runge-Lenz Vector)

(a) In notation adjusted for $d = 2$, show that the *Laplace-Runge-Lenz vector*

$$\hat{A} : \widehat{P} \to \mathbb{R}^2 \quad , \quad \hat{A}(p, q) := \hat{L}(p, q)\begin{pmatrix} p_2 \\ -p_1 \end{pmatrix} - Z\frac{q}{\|q\|}\,. \quad (11.3.7)$$

is constant in time and points in the direction of the pericenter.

(b) Show that $\|\hat{A}\|/Z = e$, where e is the eccentricity from (1.7). Prove Equation (1.7) for the conic by using $\hat{A}$. ♢

This conserved quantity, which was discovered by Jakob Hermann in 1710, permits us to regularize the collision orbits. The latter are parametrized by their energy $E \in \mathbb{R}$ and the direction $\theta \in S^{d-1}$ of the pericenter for the colliding mass point. The fact that such a direction exists at all, as a limit for almost-colliding orbits, is a consequence of the existence of A.

In geometric terms, regularizing collision orbits amounts to reflecting them in the origin of configuration space. Mathematically, regularization can be described as an extension of the dynamical system (Definition 10.5).

11.23 Theorem (Regularization of the Kepler Problem)
The Hamiltonian system $(\widehat{P}, \hat{\omega}, \hat{H})$ can be extended to a Hamiltonian system (P, ω, H) as follows:

- *As a set, the* $2d$*-dimensional differentiable manifold* P *is*

$$P = \widehat{P} \,\dot{\cup}\, \left(\mathbb{R} \times S^{d-1}\right). \tag{11.3.8}$$

- *The restriction of the smooth symplectic form* $\omega \in \Omega^2(P)$ *to* $\widehat{P}$ *is* $\hat{\omega}$.
- *The restriction of the smooth function* $H : P \to \mathbb{R}$ *to* $\widehat{P}$ *is* $\hat{H}$.
- *The Hamiltonian flow* Φ *of* (P, H, ω) *is smooth and complete, i.e.,* $\Phi \in C^\infty(\mathbb{R} \times P, P)$.

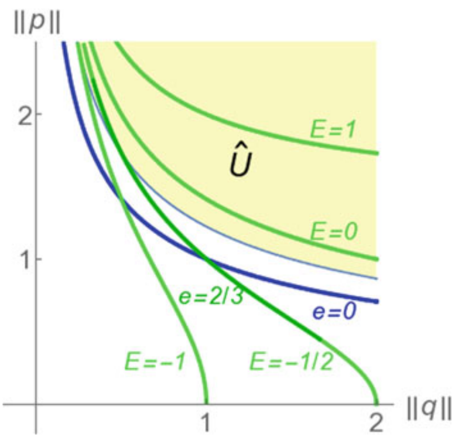
Proof:

- We investigate the Kepler dynamics in the following neighborhood $\hat{U}$ of the singularity in phase space:

$$\hat{U} := \left\{ (p, q) \in \widehat{P} \,\middle|\, \|p\|^2 > \frac{c\,Z}{\|q\|} \right\}, \tag{11.3.9}$$

where $c := \frac{3}{2}$, or any value in $(1, 2)$. Within $\hat{U}$, for an orbit $(p, q) : I \to \widehat{P}$, the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle q\,,\,p\,\rangle = \|p\|^2 - \frac{Z}{\|q\|} > \frac{c-1}{2}\frac{Z}{\|q\|} \tag{11.3.10}$$



Phase space neighborhood $\hat{U}$ of the singularity. Circular orbits (eccentricity $e = 0$) are disjoint to $\hat{U}$.
(Almost-) collision orbits ($e \sim 1$) intersect $\hat{U}$ (example: $e = 2/3$, $E = -1/2$).

is positive, so the flow $\hat{\Phi}$ there is transversal to the *pericentric hypersurface*

$$\hat{S}_0 := \left\{ (p, q) \in \widehat{P} \mid \langle q\,,\,p\,\rangle = 0 \right\}. \tag{11.3.11}$$

By $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\|q\|^2 = 2\frac{\mathrm{d}}{\mathrm{d}t}\langle q\,,\,p\,\rangle > 0$, this inequality also shows that the intersection of the orbit with $\hat{S}_0$ is indeed the pericenter, i.e., the orbit has minimal distance from the origin at the intersection point.

Every collision orbit will be within $\hat{U}$ immediately before and immediately after the collision, because an orbit with energy $\hat{H}(p, q) = E$ satisfies

$$\|p\|^2 - \frac{cZ}{\|q\|} = \frac{(2-c)Z}{\|q\|} + 2E \overset{q\to 0}{\longrightarrow} +\infty\,.$$

- By Theorem 3.39, the domain $D \subset \mathbb{R}_t \times \widehat{P}$ of the maximal flow from (11.3.4) with the upper and lower semicontinuous escape times $T^+ : \widehat{P} \to (0, \infty]$ and

$T^- : \widehat{P} \to [-\infty, 0) := \{-\infty\} \cup (-\infty, 0)$ respectively has the form

$$D = \left\{(t,x) \in \mathbb{R} \times \widehat{P} \mid t \in \big(T^-(x), T^+(x)\big)\right\}.$$

For initial values x with $T^+(x) < \infty$, there will be a collision at time $T^+(x)$, i.e.,

$$\lim_{t \nearrow T^+(x)} q(t,x) = 0\,.$$

• As noted already, the solution curves $t \mapsto q(t, x_0)$ in M with initial values $x_0 = (p_0, q_0) \in \widehat{P}$ remain in the plane (or line) through the origin that is spanned by p_0 and q_0. So let us first assume $d = 2$.
The angular momentum (11.3.5) is invariant under the flow $\hat{\Phi} : D \to \widehat{P}$.
• The direction of the Laplace-Runge-Lenz vector (written in complex coordinates)

$$\hat{\varphi} : \hat{U} \to S^1\,,\ \hat{\varphi}(p,q) := \arg\big(\hat{A}(p,q)\big) = \arg\left(-\imath p \hat{L}(p,q) - Z\frac{q}{|q|}\right) \quad (11.3.12)$$

is defined and thus smooth, because the argument of arg in (11.3.12) is not zero:

$$\left|\imath \hat{L}(x)p + Z\frac{q}{|q|}\right|^2 = 2\hat{L}(x)^2 \hat{H}(x) + Z^2 > 0 \qquad \big(x = (p,q) \in \hat{U}\big). \quad (11.3.13)$$

The inequality is obvious for $\hat{H}(x) \geq 0$. For $\hat{H}(x) < 0$, on the other hand, one concludes from $|\hat{L}(x)| \leq |p|\,|q|$ and Definition (11.3.9) of $\hat{U}$ that

$$2\hat{L}(x)^2\hat{H}(x) \;\geq\; |p|^2|q|\big(|p|^2|q| - 2Z\big) \;=\; \big(|p|^2|q| - Z\big)^2 - Z^2 \;>\; -Z^2 \quad \big(x \in \hat{U}\big).$$

For simplicity, we will denote the restrictions $\hat{f}\restriction_{\hat{U}}$ of functions $\hat{f} : \widehat{P} \to \mathbb{R}$ again by $\hat{f}$. So we have the local coordinates[7]

$$\big(\hat{H}, \hat{T}, \hat{L}, \hat{\varphi}\big) : \hat{U} \to \mathbb{R}^3 \times S^1\,. \quad (11.3.14)$$

• These coordinates are canonical in the sense of Definition 10.18, i.e., the Poisson brackets are of the form

$$\{\hat{T}, \hat{H}\} = \{\hat{\varphi}, \hat{L}\} = 1 \quad \text{and} \quad \{\hat{L}, \hat{H}\} = \{\hat{\varphi}, \hat{H}\} = \{\hat{L}, \hat{T}\} = \{\hat{\varphi}, \hat{T}\} = 0\,. \quad (11.3.15)$$

To see this, recall that the Poisson brackets with $\hat{H}$ are the derivatives $\{\hat{f}, \hat{H}\} = \frac{\mathrm{d}}{\mathrm{d}t}\hat{f} \circ \hat{\Phi}_t|_{t=0}$. Whereas $\hat{L}$ is a conserved quantity due to the potential V being centrally symmetric, see (1.4), and $\hat{\varphi}$ is also conserved according to Exercise 11.22,

[7] This wording is not precise, because the values of $\hat{\varphi}$ are not from $\mathbb{R}$, but from S^1. More precisely speaking, we equip also S^1 with coordinate charts, on which the angle is then defined as a real-valued coordinate.

$\hat{T}$ is defined as the time parameter, so its time derivative is 1. From the definition of $\hat{T}$ in (11.3.6), on can see that this quantity is invariant under rotation of position and momentum space by the same angle. As $\hat{L}$ is the Hamiltonian generating this rotation (see also Example 13.15), it follows that $\{\hat{L}, \hat{T}\} = 0$. *Mutatis mutandis*, the angle changes under the flow generated by $\hat{L}$ with speed $\{\hat{\varphi}, \hat{L}\} = 1$.

• A comparatively simple way of showing the remaining relation $\{\hat{\varphi}, \hat{T}\} = 0$ is by observing that the Hamiltonian vector field $X_{\hat{\varphi}}$ of $\hat{\varphi}$ is tangential to the hypersurface $\hat{T} \equiv 0$. To this end, we show that

$$\{\hat{\varphi},\ \langle q, p\rangle\} = 0 \tag{11.3.16}$$

on the surface $\hat{T} \equiv 0$. This surface in phase space equals $\hat{S}_0 \cap \hat{U}$ with $\hat{S}_0$ from (11.3.11). On this surface, both complex numbers showing up in Definition (11.3.12) of $\hat{\varphi}$ have the same argument *modulo* π, and this argument is invariant under the dilation $(p, q) \mapsto (e^{-t} p, e^t q)$ generated by $\langle q, p\rangle$, hence (11.3.16). On the other hand, $\{\hat{\varphi}, \hat{T}\}$ is invariant under the flow $\hat{\Phi}_t$ of $\hat{H}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\Phi}_t^*\big(\{\hat{\varphi}, \hat{T}\}\big) = -\hat{\Phi}_t^*\big(\{\hat{H}, \{\hat{\varphi}, \hat{T}\}\}\big) \;=\; \hat{\Phi}_t^*\Big(\big\{\hat{\varphi}, \{\hat{T}, \hat{H}\}\big\} + \big\{\hat{T}, \{\hat{H}, \hat{\varphi}\}\big\}\Big) = 0,$$

where we have used the Jacobi identity (E.21). Except for collision orbits, all orbits in $\hat{U}$ intersect the hypersurface $\hat{S}_0$. As the collision orbits, which are characterized by vanishing angular momentum, form a subset that is nowhere dense in $\hat{U}$, the equality $\{\hat{\varphi}, \hat{T}\} = 0$ holds everywhere.

• In the chart (11.3.14), the incomplete flow $\hat{\Phi}$ on $\hat{U}$ generated by $\hat{H}$ has therefore been linearized.

The points $(p, q) \in \hat{U}$ on collision orbits have angular momentum $\ell = 0$, but $\hat{T}(p, q) \neq 0$. The cylinder $\mathbb{R} \times S^{d-1}$ in the set $P = \widehat{P} \,\dot{\cup}\, (\mathbb{R} \times S^{d-1})$ from (11.3.8) will then be identified with the set of missing phase space points that is characterized by $(\ell, t) = (0, 0)$. Then P obtains the structure of a differentiable manifold by introducing a chart on

$$U \;:=\; \hat{U} \,\dot{\cup}\, \big(\mathbb{R} \times S^{d-1}\big) \;\subset\; P\,,$$

initially for the case $d = 2$. Namely we extend (11.3.14) to a mapping

$$\big(H, T, L, \varphi\big) : U \to \mathbb{R}^3 \times S^1 \tag{11.3.17}$$

by defining for $(E, \theta) \in \mathbb{R} \times S^1$:

$$\big(H, T, L, \varphi\big)(E, \theta) := (E, 0, 0, \theta)\,.$$

• From (11.3.15), we obtain the identity

$$\hat{\omega}\restriction_{\hat{U}} = d\hat{T} \wedge d\hat{H} + d\hat{\varphi} \wedge d\hat{L}$$

for the restrictions $\hat{\omega} \in \Omega^2(\widehat{P})$ of the canonical symplectic form ω_0. Therefore, we define conversely the extension $\omega \in \Omega^2(P)$ of $\hat{\omega}$ by

$$\omega\restriction_U := dT \wedge dH + d\varphi \wedge dL\,.$$

• Since ω is symplectic, and H, viewed as an extension of $\hat{H} : \widehat{P} \to \mathbb{R}$ defined on all of P by (11.3.17), is smooth, we can use H as Hamiltonian. Its flow Φ is complete, and is linear in the local coordinates (11.3.17).
• For $d = 2$ degrees of freedom, the values (ℓ, θ) of angular momentum and pericenter angle are points on the cylinder T^*S^1; analogously, for arbitrary d and $(p, q) \in \hat{S}_0$, the values of the "angular momentum vector" $\|q\|\, p$ and the pericenter direction $q/\|q\|$ are elements of T^*S^{d-1}. This allows the generalization of the above construction to d degrees of freedom. □

11.24 Remarks (Kepler Problem)

1. Whereas the motion of two mass points under gravitational attraction is singular in (p, q)-coordinates, it is smooth in adapted coordinates.
2. Just as every central force problem is integrable, see Section 13.1, so is the Kepler problem in particular. But similar to the harmonic oscillator in d degrees of freedom, and with equal frequencies, as analyzed in Theorem 6.35, it has more than d independent constants of motion (namely $2d - 1$). Such Hamiltonian systems are called *superintegrable*. They enjoy special properties. In the examples mentioned here, the periodicity of all orbits on compact energy surfaces follows.
3. The method of regularization presented here has the advantage over other variants that the phase space gets extended, rather than changed altogether, that the time parametrization remains the same, and that the flow gets completed on all energy surfaces at the same time.
   On the other hand, its geometric content is less transparent than for instance in the case of regularization by the *Levi-Civita transformation*: This regularization relies on the observation that the holomorphic mapping
   $$\mathbb{C} \longrightarrow \mathbb{C} \quad , \quad Q \longmapsto q := Q^2$$
   maps straight lines to parabolas (see Figure 11.3.1), and more generally the conics of the Kepler orbits are images of conics that are regular in the origin. The corresponding local point transformation (see (10.32)) of the phase space transforms the momenta according to $p = \frac{P}{2\overline{Q}}$, and thus after reparametrization of time, it leads to the motion in the potential of a harmonic oscillator.
   The analog of the Levi-Civita transformation for three degrees of freedom is called the *Kustaanheimo-Stiefel transformation*, or *Cayley-Klein parametrization* (and is based on the Hopf mapping (6.36)), see STIEFEL and SCHEIFELE [StSc].
4. Another method of regularization is described in the article [Mos3] by MOSER. It shows that for energies $E < 0$, the restricted flow Φ_t $(t \in \mathbb{R})$ on the energy

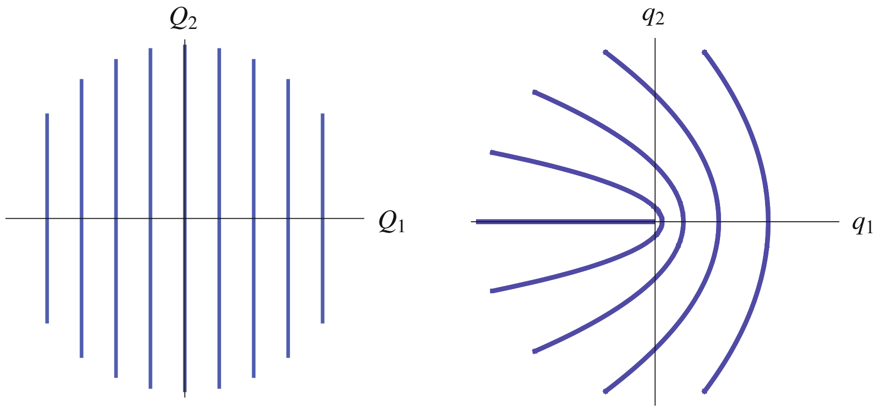


Figure 11.3.1 The effect of the Levi-Civita transformation $\mathbb{C} \to \mathbb{C}$, $Q \mapsto Q^2$, i.e., $(q_1, q_2) = (Q_1^2 - Q_2^2, 2Q_1 Q_2)$ on straight lines in the Q-plane

surface $\Sigma_E := H^{-1}(E) \subset P$ is conjugate, in the sense of Definition 2.28, to the geodesic flow

$$\Psi_t : T_1 S^d \to T_1 S^d \quad , \quad \Psi_t(x, y) = \big(\cos(t)x + \sin(t)y, -\sin(t)x + \cos(t)y\big)$$

$(t \in \mathbb{R})$ on the unit tangent bundle

$$T_1 S^d = \big\{(x, y) \in \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \mid \|x\| = \|y\| = 1\,,\ \langle x, y\rangle = 0\big\} \qquad (11.3.18)$$

of the d-dimensional sphere. In particular, Σ_E is diffeomorphic to $T_1 S^d$.
More specifically, the sphere minus its north pole n is projected stereographically to the momentum plane $\mathbb{R}_p^d$ of the phase space $\widehat{P} \cong \mathbb{R}_p^d \times \widehat{M}$ of the unregularized Kepler problem, and the unit tangent bundle is mapped under the linearized map to $\hat{H}^{-1}(E) \subset \widehat{P}$. So the sphere of collisions S^{d-1} of the Kepler problem corresponds to $\{(x, y) \in T_1 S^d \mid x = n\}$.
This transformation also shows that the Kepler problem has special symmetries. For, whereas general central potentials (as discussed in more detail in Example 13.15) are invariant under the action of the rotation group $\mathrm{SO}(d)$, the larger group $\mathrm{SO}(d + 1)$ operates on the unit tangent bundle (11.3.18). ♢

11.25 Exercises (Regularizable Singular Potentials)

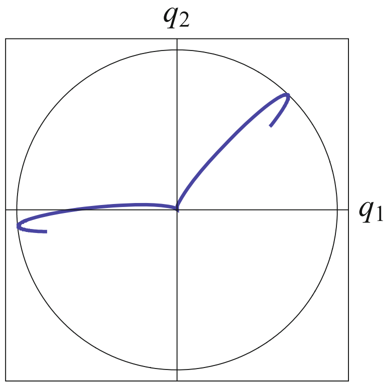
1. For which homogeneous singular potentials of the form $q \mapsto -Z/\|q\|^a$ with $a \in (0, 2)$ is it possible to regularize the Hamiltonian dynamics analogous to the Kepler case $a = 1$?

Hint: Consider the total deflection angle $\Delta\varphi$ for energy $E > 0$

$$\Delta\varphi(E,\ell) = 2\int_{r_{\min}}^{\infty} \frac{\dot\varphi}{\dot r}\,\mathrm{d}r = 2\int_{r_{\min}}^{\infty} \frac{\ell/r^2}{\sqrt{2(E-V_\ell(r))}}\,\mathrm{d}r$$

with an effective potential $V_\ell(r) := \frac{\ell^2}{2r^2} - \frac{Z}{r^a}$ in the limit of vanishing angular momentum ℓ. Show that this limit $\lim_{\pm\ell\searrow 0}\Delta\varphi(E,\ell)$ is independent of the energy, namely it equals

$$= \pm 2\int_1^{\infty} \frac{\mathrm{d}u}{u\sqrt{u^{2-a}-1}} = \pm\frac{2\pi}{2-a}\,.$$



Almost collision orbit with negative energy in a non-regularizable potential ($V(q) = -1/\|q\|^{1.3}$)

2. Conclude from Remark 11.24.4 that for negative energies E and d degrees of freedom and the energy shell Σ_E of the Kepler problem, the space Σ_E/S^1 of orbits
   (a) is diffeomorphic to S^2 when $d = 2$,
   (b) is diffeomorphic to $S^2 \times S^2$ when $d = 3$.
   (c) Show directly for the Kepler problem with $d = 2$ degrees of freedom that $\Sigma_E/S^1 \cong S^2$, when $E < 0$.
   What is the shape of this orbit space when $E > 0$?

Hints:

- For $d = 2$ and $(x, y) \in T_1S^2$, consider the vector $x \times y \in \mathbb{R}^3$.
- For $d = 3$, identify the vector space $\mathbb{R}^4 \times \mathbb{R}^4$ from Definition (11.3.18) of T_1S^3 with the cartesian product $\mathbb{H} \times \mathbb{H}$ of the skew field of quaternions $\mathbb{H}$ from E.27, and show that the mapping

$$T_1S^3/S^1 \to S^2 \times S^2 \subset \mathrm{Im}\mathbb{H} \times \mathrm{Im}\mathbb{H} \quad , \quad [(x,y)] \mapsto \big(xy^*, y^*x\big)$$

  (where $[(x, y)]$ is the orbit through (x, y)) is well-defined and is a diffeomorphism.
- Use the Laplace-Runge-Lenz vector $A : P \to \mathbb{R}^2$ obtained from $\hat A$ in (11.3.7) by regularization. Show that $\|A\|^2 = 2\|L\|^2 H + Z^2$, hence for energy $E < 0$ and $x \in \Sigma_E$, the values of A lie in the disc $\|A(x)\| \le Z$. How many orbits in Σ_E correspond to a value a of A, satisfying $\|a\| = Z$ or $\|a\| < Z$ respectively? ◊

The form of the orbit space obtained in Exercise 11.25.2 is essential for perturbation theory. It is used in [Mos3] by MOSER to show that for small perturbations of the potential $\hat H$, at least 3 periodic orbits of a given energy $E < 0$ survive when $d = 2$, and likewise at least 6, when $d = 3$.

11.3.2 Two Centers of Gravitation

It is only in the case $n = 1$, which was just considered, that the *n-center problem* can be viewed as a special case of an n-body problem. We will however see that the two-center problem has nevertheless applications in celestial mechanics.

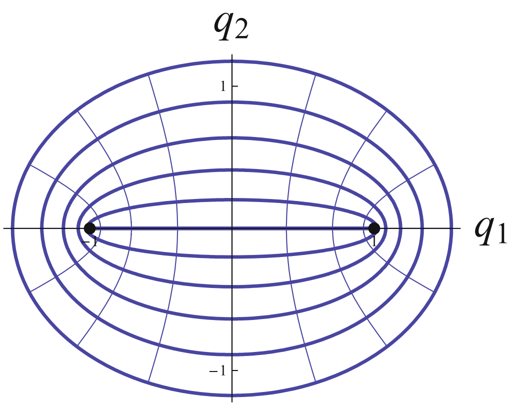
In the n-center problem, one studies the motion of a mass point in the gravitational field of gravitational centers fixed in n different locations $s_1, \ldots, s_n \in \mathbb{R}^d$. This makes the problem simpler than the $(n+1)$-body problem.

So for the case $d = 3$ and $n = 2$ that we are about to consider, the configuration space is $\widehat{M} := \mathbb{R}^3 \setminus \{s_1, s_2\}$, and for masses or charges $Z_1, Z_2 \in \mathbb{R} \setminus \{0\}$ respectively, the function

$$V : \widehat{M} \to \mathbb{R} \quad , \quad V(q) = \frac{-Z_1}{\|q - s_1\|} + \frac{-Z_2}{\|q - s_2\|}$$

is a smooth potential. While for $Z_k > 0$, the Hamiltonian $\hat{H} : \widehat{P} \to \mathbb{R}$, $\hat{H}(p, q) = \frac{1}{2}\|p\|^2 + V(q)$ on the phase space $\widehat{P} := T^*\widehat{M} \cong \mathbb{R}^d \times \widehat{M}$ generates an incomplete dynamics, the regularization at s_k can be performed exactly as in the 1-center problem of Sect. 11.3.1.

Amazingly, this system is integrable. This was shown by Jacobi, see [Jac]. Subsequently, the dynamics was repeatedly investigated in more detail. Waalkens, Dullin, and Richter [WDR] contains an extensive literature list on the 2-center problem. By a Euclidean motion and a rescaling, we may restrict our discussion to the special case $s_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $s_2 := \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$. Specifically, the integration is performed by means of the use of prolate spheroidal coordinates $(\xi, \eta, \varphi) \in (0, \infty) \times (0, \pi) \times S^1$ for $\widehat{M}$; see also Chapter 4.3 of Thirring [Th1]: Here



$$q \equiv \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = F(\xi, \eta, \varphi) := \begin{pmatrix} \cosh(\xi)\cos(\eta) \\ \sinh(\xi)\sin(\eta)\cos(\varphi) \\ \sinh(\xi)\sin(\eta)\sin(\varphi) \end{pmatrix}.$$

Level sets of ξ and η are depicted in the figure. The Jacobi determinant is positive on the entire domain $(0, \infty) \times (0, \pi) \times S^1$:

$$\det\bigl(DF(\xi, \eta, \varphi)\bigr) = \bigl(\cosh^2(\xi) - \cos^2(\eta)\bigr)\sinh(\xi)\sin(\eta)\,; \qquad (11.3.19)$$

but it vanishes on the axis through s_1 and s_2 (because this q_1-axis is pointwise invariant under rotation by an angle $\varphi \in S^1$). The prolate spheroidal coordinates are well adapted to the two center problem, because the distances to the centers are of the simple form[8]

$$\|q - s_1\| = \cosh(\xi) - \cos(\eta) \quad , \quad \|q - s_2\| = \cosh(\xi) + \cos(\eta)\,.$$

So, letting $Z_\pm := Z_2 \pm Z_1$ and $f(\xi,\eta) := \cosh^2(\xi) - \cos^2(\eta)$, the potential is

$$V \circ F(\xi,\eta,\varphi) = \frac{-Z_+\cosh(\xi) + Z_-\cos(\eta)}{f(\xi,\eta)}\,.$$

For the Hamiltonian equations, we still need momenta $(p_\xi, p_\eta, p_\varphi)$ that are adjusted to the coordinate system (ξ, η, φ). As the velocities are transformed by the Jacobi matrix $\mathrm{D}F$, the dual momenta are defined by the relation

$$p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \left(\mathrm{D}F(\xi,\eta,\varphi)^{-1}\right)^\top \begin{pmatrix} p_\xi \\ p_\eta \\ p_\varphi \end{pmatrix}.$$

Therefore

$$p_\varphi = \frac{\partial F_1}{\partial\varphi} p_1 + \frac{\partial F_2}{\partial\varphi} p_2 + \frac{\partial F_3}{\partial\varphi} p_3 = q_2 p_3 - q_3 p_2$$

is a constant of motion: Letting $W(q_1, r) := \frac{-Z_1}{\sqrt{(q_1-1)^2+r^2}} + \frac{-Z_2}{\sqrt{(q_1+1)^2+r^2}}$ and $r := \sqrt{q_2^2+q_3^2}$, one has $V(q) = W(q_1, r)$, and therefore

$$\frac{\mathrm{d}p_\varphi}{\mathrm{d}t} = \dot q_2 p_3 - \dot q_3 p_2 + q_2 \dot p_3 - q_3 \dot p_2 = p_2 p_3 - p_3 p_2 + \frac{\partial W}{\partial r}(q_1, r)\,\left(q_2 \frac{q_3}{r} - q_3 \frac{q_2}{r}\right) = 0.$$

p_φ is the first component of the angular momentum vector. By the Noether theorem (Theorem 13.22), the constancy of p_φ follows geometrically from the invariance of V under rotations about the q_1-axis.

The kinetic energy has the form

$$\tfrac12 \|p\|^2 = \tfrac12 (p_\xi, p_\eta, p_\varphi)\left(\mathrm{D}F(\xi,\eta,\varphi)^{-1}\right)\left(\mathrm{D}F(\xi,\eta,\varphi)^{-1}\right)^\top \begin{pmatrix} p_\xi \\ p_\eta \\ p_\varphi \end{pmatrix}.$$

For initial values $x_0 \in \widehat{P}$, let $\ell := p_\varphi(x_0)$ and $E := \hat H(x_0)$. Then, using $\left(\mathrm{D}F^{-1}\right)\left(\mathrm{D}F^{-1}\right)^\top = \left(\mathrm{D}F^\top \mathrm{D}F\right)^{-1}$ and

$$\mathrm{D}F^\top \mathrm{D}F\ (\xi,\eta,\varphi) = \frac{\mathrm{diag}(1, 1, \sinh(\xi)^{-2} + \sin(\eta)^{-2})}{f(\xi,\eta)}\,,$$

one obtains $\hat H - E$ in the new coordinates as

[8] As these relations imply $\cosh(\xi) = \frac12(\|q - s_1\| + \|q + s_1\|)$, the level sets of ξ are indeed rotation ellipsoids with foci s_i.

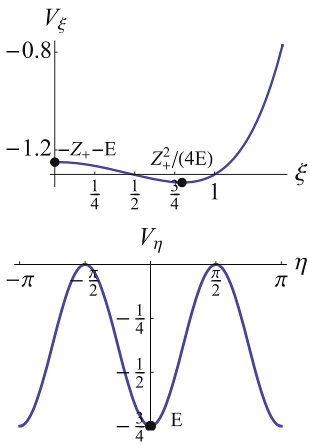
$$f(\xi,\eta)^{-1}\big(H_\xi(p_\xi,\xi)+H_\eta(p_\eta,\eta)\big),$$

where $H_\xi(p_\xi,\xi) := \frac{1}{2}p_\xi^2 + V_\xi(\xi)$ with

$$V_\xi(\xi) := \frac{\ell^2}{2\sinh^2(\xi)} - Z_+\cosh(\xi) - E\cosh^2(\xi)$$

and $H_\eta(p_\eta,\eta) := \frac{1}{2}p_\eta^2 + V_\eta(\eta)$ with

$$V_\eta(\eta) := \frac{\ell^2}{2\sin^2(\eta)} + Z_-\cos(\eta) + E\cos^2(\eta)\,.$$



The potentials V_ξ and V_η for $\ell = 0$, $Z_1 = Z_2 = 1$ and $E = -3/4$

Transitioning to the extended phase space $\widehat{P} \times T^*\mathbb{R} \cong \widehat{P} \times \mathbb{R}_E \times \mathbb{R}_s$ and using the new time parameter s defined by

$$\frac{\mathrm{d}t}{\mathrm{d}s} = f(\xi,\eta)\,,$$

the new Hamiltonian[9] $\mathcal{H} : \widehat{P} \times T^*\mathbb{R} \to \mathbb{R}$ given by:

$$\mathcal{H}(p_\xi,\xi,p_\eta,\eta;E,s) = f(\xi,\eta)(\hat{H} - E) = H_\xi(p_\xi,\xi) + H_\eta(p_\eta,\eta)$$

separates. The motion on $\mathcal{H}^{-1}(0)$ coincides—up to the parametrization by time—with the motion on $\hat{H}^{-1}(E)$. For

$$K := H_\xi(x_0) = -H_\eta(x_0)\,,$$

we have three constants of motion, which are in general independent of each other, namely $\hat{H}$, H_ξ, and p_ϕ; and their values are E, K, and ℓ.

In Figure 11.3.2, we see three trajectories in the q_1-q_2-plane (for which the angular momentum is therefore $\ell = 0$). They lie entirely in the Hill domain defined by $V(q) \le E$. This domain is symmetric under reflection in the q_1-axis (and in this case also the q_2-axis because we have chosen $Z_1 = Z_2 = 1$). Moreover, it is connected, because the value $E = -3/4$ of the total energy surpasses the value $V(0) = -Z_+ = -Z_1 - Z_2$ of the saddle point. We observe the following:

- In the figure on the left, the projected 2-torus does not contain any of the singularities s_1, s_2; rather the trajectories move around these singularities, say, counterclockwise. By time reversibility $(p_1,p_2,q_1,q_2) \mapsto (-p_1,-p_2,q_1,q_2)$, there is a second invariant torus with the same constants K and E, which corresponds to a clockwise motion.

[9]This definition of $\mathcal{H}$ is written a bit shoddily, because $\hat{H}$ is a function of the *cartesian* coordinates.

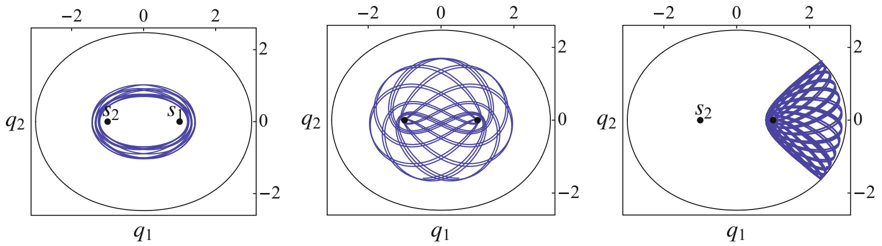


Figure 11.3.2 Three trajectories in the two center potential for energy $E = -3/4$, angular momentum $\ell = 0$ and various choices K_i for the constants of motion H_ξ

- In the figure in the middle, the projection of the invariant 2-torus on the configuration space contains both singularities s_1, s_2.
- In the figure on the right, the trajectory remains near s_1. For the value K of the constants of motion, there exists (due to the symmetry $(p_1, p_2, q_1, q_2) \mapsto (-p_1, p_2, -q_1, q_2)$) a second 2-torus, along with the 2-torus of the depicted trajectory; its projection onto the Hill domain contains only s_2.

The *bifurcation set* is given as the set of those values for which the mapping from the phase space to the space of constants of motion is not *locally trivial* (see Definition 7.23). For the case $\ell = 0$, we obtain planar motion.

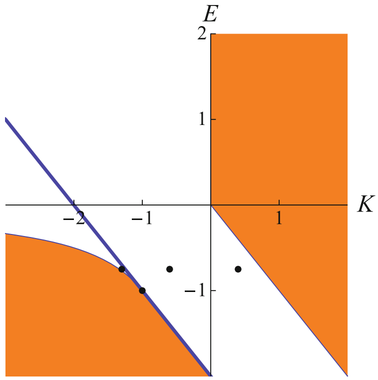
Investigating the extrema of V_ξ and V_η, we observe that for the the simplest (2-dimensional) case $\ell = 0$, the image of $(\hat{H}, H_\xi)$ is the closure of a domain in $\mathbb{R}^2$ that is delimited by the graphs of the following curves: From $K = H_\xi \geq V_\xi$, one obtains $K \geq K_+(E)$ with

$$K_+(E) := \begin{cases} \frac{Z_+^2}{4E} & , 0 > E > -\frac{Z_+}{2} \\ -(Z_+ + E), & E \leq -Z_+/2 \end{cases};$$

from $-K = H_\eta \geq V_\eta$, one obtains $K \leq K_-(E)$ with

$$K_-(E) := \begin{cases} \frac{Z_-^2}{4E} & , E > |Z_-|/2 \\ |Z_-| - E & , E \leq |Z_-|/2 \end{cases}.$$

For the symmetric situation $Z_1 = Z_2$, the bifurcation diagram (see figure) is the union of the two boundary curves $K_\pm$ and the three straight lines in the image of $(\hat{H}, H_\xi)$.



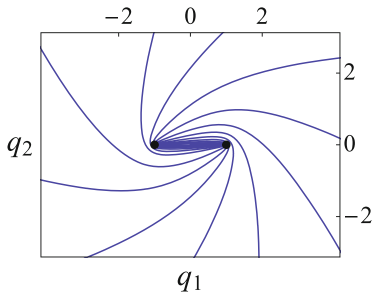
Bifurcation diagram for the two center problem with $\ell = 0$ and $Z_1 = Z_2 = 1$

- $E = 0$ corresponds to the transition from compact to non-compact energy surfaces,
- $K = 0$ corresponds to the critical value of V_η at $\eta = \pm\pi/2$

- and $K_0(E) := -(Z_+ + E)$. This line K_0 corresponds to the $(\hat{H}, H_\xi)$-values of the periodic orbit that oscillates between s_1 and s_2, hence has constant coordinate $\xi = 0$.

In the bifurcation diagram, the three points with energy $E = -3/4$ correspond in their position (left, middle, right) to the three trajectories from Figure 11.3.2. In WAALKENS, DULLIN and RICHTER [WDR], the bifurcation is also analyzed for the case of unequal masses/charges and for $d = 3$ degrees of freedom.

Not all trajectories of positive energy E are scattering orbits. The bound orbit that shuttles between the two centers is hyperbolic and thus has stable and unstable manifolds. Both of them have dimension d and have the parameter value $K(E)$. In the figure on the right, one can see for $d = 2$ the projections on the configuration space of a family of trajectories that form such a manifold.



Due to the reversibility of the motion, these projected orbits look alike for *escape orbits*, which are bound in the past, and *capture orbits*, which are bound in the future. Only the direction in which they are traversed is different.

11.26 Remark (The Gravitation Field of the Earth)
In 1687, Newton noticed that the rotating earth should be flattened at its poles. In *Proposition XIX, Problem III* of volume III of his *Principia* [Ne], he inferred from a calculation that treated the earth like a drop of liquid:

> "…and therefore the diameter of the earth at the equator is to its diameter from pole to pole as 230 to 229."

Essentially, this was confirmed by measurements.[10] Even though the surface of the earth with its mountains features differences in elevation by several kilometers, the *geoid*, which is the idealized equipotential surface of the gravitational field at sea level, differs from a rotational ellipsoid by barely 100 metres.

This rotational ellipsoid is almost fixed in space. Therefore satellite orbits are integrable in good approximation. Whereas the two center model models a prolate (cigar shaped) rotation ellipsoid, it can be applied, by analytic continuation, to the oblate (flattened) shape of the earth.

Nowadays, satellite orbits can be measured to centimeter precision by GPS and other techniques. This allows to determine deviations of the (harmonic) gravitational field from the field predicted by the two center model with high precision. While this

[10]The actual flattening is only about 1/298. As the density of the earth increases towards its center, Newton's calculation only gives an upper bound for the flattening.
Conversely, HUYGENS in [Huy] assumed that the mass of the earth is concentrated in its center. From the requirement that the sum of gravitation and centrifugal force be normal to the surface of the earth, he calculated a flattening of $1/578$—which is a lower bound.

does not immediately determine the density distribution in the earth (see Exercise 12.37 in this context), important conclusions about it can nevertheless be drawn. ♢

11.3.3 The n-Body Problem

"D'ailleurs, ce qui nous rend ces solutions périodiques si précieuses, c'est qu'elles sont, pour ainsi dire, la seule bréche par où nous puissions essayer de pénétrer dans une place jusqu'ici réputée inabordable." HENRI POINCARÉ, in: Les Méthodes nouvelles de la mécanique céleste, Volume 1, §36.[11]

In contradistinction to the two-body problem, the motion of three or more mass points is not integrable. The best mathematicians since the days of Newton attempted to find an explicit solution for the equations of motion. But it became understood only gradually that the failure of these attempts is due to properties of the system itself, and not due to insufficient human skills.

In 1887, HEINRICH BRUNS showed in [Br] that next to the known 10 integrals of motion (Exercise 11.21), the n-body problem has no further independent integrals that are algebraic functions on the phase space. This alone does not imply that the system isn't integrable, but it showed that certain approaches at a solution would have to be doomed to failure (see also [Whi] by WHITTAKER, §164).

But already in the 18th century, certain special solutions to the n-body problem were found, solutions that had a high symmetry. To this end, a study within the center of mass system, i.e., on the manifold

$$\widehat{M}_0 := \left\{q \in \widehat{M} \;\middle|\; \textstyle\sum_{k=1}^n m_k q_k = 0\right\}.$$

is sufficient. The diagonal matrix $\mathcal{M} := \bigoplus_{k=1}^n m_k \mathbb{1}_d \in \mathrm{Mat}(nd, \mathbb{R})$ is called the *mass matrix*.[12] Using this matrix, the Hamiltonian equations of (11.3.1) can be written in the form

$$\mathcal{M}\dot{q} = p\,,\;\; \dot{p} = -\nabla V(q) \qquad \text{, or} \qquad \mathcal{M}\ddot{q} = -\nabla V(q)\,.$$

11.27 Examples (Central Collisions)
Let us begin by deriving special solutions of the form $q(t) = r(t)q_0$ with $q_0 \in \widehat{M}_0$ and $r(t) > 0$, $r(0) = 1$. As ∇V is homogenous of degree -2, it follows that

$$\ddot{r}\,\mathcal{M}q_0 = -r^{-2}\,\nabla V(q_0) \qquad \text{, hence} \quad c\,\mathcal{M}q_0 = \nabla V(q_0) \text{ and } \ddot{r} = -cr^{-2}\,,$$

[11] Translation: "Moreover, what makes these periodic solutions so valuable for us is that they are, if we may say so, the only breach by which we can try to enter into a location that was until now considered to be unreachable." Indeed, the semiclassical theory of chaotic systems is successfully based on periodic orbits. Keywords in this context are the trace formulas by Selberg and by Gutzwiller. This is explained in detail in the internet based *Chaos Book* [Cv] by CVITANOVIĆ and others.

[12] In (12.6.2) we define the scalar product corresponding to $\mathcal{M}$.

with $c > 0$. We already know from Exercise 3.41 that the Hamiltonian differential equation $\ddot{r} = -cr^{-2}$ only has solutions that feature collisions in at least one direction in time. As all particles collide simultaneously in the origin, we are talking about a *central collision*.

For every solution of $c\,\mathcal{M}q_0 = \nabla V(q_0)$, the constant of proportionality is $c = \frac{-V(q_0)}{\langle q_0, \mathcal{M}q_0\rangle}$. This follows by scalar multiplication of both sides with q_0, because $\langle q_0, \nabla V(q_0)\rangle = -V(q_0)$.

The equation $c\,\mathcal{M}q_0 = \nabla V(q_0)$ is

- solvable for every $q_0 \in \widehat{M}_0$, when there are $n = 2$ particles, because in the center of mass system the two products 'mass times location' for either particle are each other's negative.
- In the case of $n \geq 3$ particles, one can still construct new solutions from a given solution q_0, by rotation or reflection with a matrix $O \in \mathrm{O}(d)$, and by dilation with a factor $\lambda > 0$; the new solution being then $\lambda(O \oplus \ldots \oplus O)q_0$. ◊

The above ansatz can also be generalized to find solutions to the n-body problem that do not have collisions. The following notions are customary:

11.28 Definition

- *A configuration $q \in \widehat{M}_0$ is called* **central** *if the vectors $\nabla V(q)$ and $\mathcal{M}q$ are linearly dependent.*
- *Two central configurations $q, q' \in \widehat{M}_0$ are called* **equivalent** *if they are transformed into each other by orthogonal transformations from* O(d) *and dilations.*
- *A solution $I \ni t \mapsto q(t) \in \widehat{M}_0$ of the n-body problem is called* **homographic** *if for all times $s, t \in I$ the configurations $q(s)$ and $q(t)$ are equivalent.*
- *A configuration $q = (q_1, \ldots, q_n) \in \widehat{M}_0$ is called* **collinear** *(resp.* **planar***) if* $\mathrm{span}(q_1, \ldots, q_n) \subset \mathbb{R}^d$ *is one dimensional (resp. at most two dimensional).*

11.29 Theorem *Let $\hat{q} = (q_1, \ldots, q_n) \in \widehat{M}_0$ be a planar central configuration. So without loss of generality $d = 2$. Let $t \mapsto \big(r(t), \varphi(t)\big)$ be a solution to the Kepler problem (11.3.3) with parameter $Z := \frac{V(\hat{q})}{\langle \hat{q}, \mathcal{M}\hat{q}\rangle}$, in polar coordinates with* $r(0) = 1$. *Then, for the matrix function $t \mapsto A(t) := r(t)R(t)$ with $R(t) := \begin{pmatrix} \cos(\varphi(t)) & -\sin(\varphi(t)) \\ \sin(\varphi(t)) & \cos(\varphi(t)) \end{pmatrix}$,*

$$t \mapsto q(t) := \big(A(t)q_1, \ldots, A(t)q_n\big) \in \widehat{M}_0$$

is a homographic solution to the n-body problem.

Proof: Letting $\mathbb{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, we have

$$\begin{aligned}\ddot{q}(t) = &\big(\ddot{r}(t) - r(t)\dot{\varphi}^2(t)\big)\,\big(R(t)q_1, \ldots, R(t)q_n\big) \\ &+ \big(r(t)\ddot{\varphi}(t) + 2\dot{r}(t)\dot{\varphi}(t)\big)\big(R(t)\mathbb{J}q_1, \ldots, R(t)\mathbb{J}q_n\big)\,.\end{aligned}$$

Due to the form (1.5) of the Kepler DE in polar coordinates, the second term in the sum drops out. Due to Newton's equation and the fact that $\hat{q}$ is a central configuration, we conclude $\ddot{q} = -\mathcal{M}^{-1}\nabla V(q) = \frac{V(q)}{\langle q, \mathcal{M}q\rangle} q = \frac{V(\hat{q})}{\langle \hat{q}, \mathcal{M}\hat{q}\rangle}\frac{q}{r^3}$. This is also compatible with the radial component of Kepler's equation (1.5), when Z is the constant given above. □

So we are looking for central configurations of the n-body problem with $n \geq 2$ that are not equivalent.

11.30 Theorem (**Moulton** [Mou])
There are exactly $n!/2$ equivalence classes of collinear central configurations.

Proof: [Partly following MOECKEL[13]]

- Without loss of generality, we assume for representatives $q \in \widehat{M}_0$ of the equivalence classes that

$$\mathrm{span}(q_1, \ldots, q_n) = \mathrm{span}(e_1) \qquad \big(\text{with } e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^d\big).$$

  Accordingly, we write $q_k = r_k e_1$. So there is exactly one permutation $\pi \in S_n$ for which $r_{\pi(1)} < r_{\pi(2)} < \cdots < r_{\pi(n)}$. We now show conversely that for every $\pi \in S_n$, there exists exactly one sequence $r = (r_1, \ldots, r_n)$ satisfying $r_{\pi(1)} < r_{\pi(2)} < \cdots < r_{\pi(n)}$, $\sum_{k=1}^n m_k r_k = 0$, and the normalization $\sum_{k=1}^n m_k r_k^2 = 1$ that satisfies the scalar analog of the equation of motion $c\,\mathcal{M}q = \nabla V(q)$, namely

$$F(r) := \tilde{\mathcal{M}}^{-1}\nabla U(r) + U(r)r = 0 \quad \text{with} \quad U(r) := -\sum_{1\leq k<\ell\leq n} \frac{m_k m_\ell}{|r_k - r_\ell|} \tag{11.3.20}$$

  where $\tilde{\mathcal{M}} := \mathrm{diag}(m_1, \ldots, m_n)$. (Note that in the normalization $\sum_{k=1}^n m_k r_k^2 = 1$, one has $c = -U(r)$.)

- With this $r = (r_1, \ldots, r_n)$, we associate a representative $(q_1, \ldots, q_n)$ of an equivalence class of collinear central configurations.
  Solutions for $\pi \neq \sigma \in S_n$ will correspond to the same equivalence class if and only if $\pi(k) = \sigma(n-k)$ $(k = 1, \ldots, n)$, i.e., if and only if the order is reversed along the axis. The claimed number of $n!/2$ equivalence classes follows.

- The force F from (11.3.20) is tangential to the sphere

$$S_{\tilde{\mathcal{M}}} := \left\{r \in \mathbb{R}^n \mid \textstyle\sum_{k=1}^n m_k r_k = 0,\ \langle r, \tilde{\mathcal{M}}r\rangle = 1\right\},$$

  because the vector $\tilde{\mathcal{M}}r$, which is orthogonal to the sphere at r, has the scalar product

$$\langle \tilde{\mathcal{M}}r, F(r)\rangle = \langle r, \nabla U(r)\rangle + \langle r, \tilde{\mathcal{M}}r\rangle U(r) = \langle r, \nabla U(r)\rangle + U(r) = 0\,.$$

[13] R. Moeckel, Celestial Mechanics (Especially Central Configurations), Course at CIME, Trieste (1994).

- Now we study, without loss of generality, the domain $G := \{r \in S_{\tilde{\mathcal{M}}} \mid r_1 < \cdots < r_n\}$ on the sphere that corresponds to the identical permutation. As the intersection of $S_{\tilde{\mathcal{M}}}$ with the $n-1$ half spaces $r_k < r_{k+1}$, this domain G is diffeomorphic to an open $(n-2)$-simplex, when $n > 2$. The restriction $U_G : G \to \mathbb{R}$ of the potential U to G is smooth. Since by definition, the tangent vectors $v \in T_r G$ are perpendicular to the vector r with respect to the scalar product $\langle \cdot, \cdot \rangle_{\tilde{\mathcal{M}}} := \left\langle \cdot, \tilde{\mathcal{M}} \cdot \right\rangle$, it follows that

$$\langle F(r), v \rangle_{\tilde{\mathcal{M}}} = \left\langle \tilde{\mathcal{M}}^{-1} \nabla U(r) + U(r) r, v \right\rangle_{\tilde{\mathcal{M}}} = \langle \nabla U(r), v \rangle = \mathrm{D}U(r)\, v\,.$$

  So F is the gradient vector field of U_G with respect to the scalar product $\langle \cdot, \cdot \rangle_{\tilde{\mathcal{M}}}$. In particular, the location of a maximum $r \in G$ for U_G satisfies the condition $F(r) = 0$.

- Such a maximum actually exists, because $\lim_{r \to \partial G} U_G(r) = -\infty$.

- The location of the maximum is unique in G, because U_G is a strictly concave function (show this as an exercise). □

So for the 3-body problem, one obtains 3 families of collinear central configurations. They were already found by EULER in 1767 in [Eu].

In 1772, LAGRANGE [Lag] found another family of planar central configurations to the 3-body problem. In the astronomy literature, this family is split up into two families. Given the positions of the first two mass points, there are two points (called the *Lagrange points* L_4 and L_5) for the third mass point that form an equilateral triangle with the other two mass points. In comparison, the three locations for the third mass point in the collinear Euler solutions, are denoted as L_1, L_2 and L_3.

11.31 Theorem (Lagrange) *The non-collinear central configurations of the 3-body problem have the shape of an equilateral triangle.*

Proof: The central configuration $q = (q_1, q_2, q_3) \in \widehat{M}_0$ satisfies the equations

$$c\, q_k = \sum_{\ell \neq k} m_\ell \frac{q_k - q_\ell}{\|q_k - q_\ell\|^3} \qquad (k = 1, 2, 3).$$

By summation, one obtains, with the help of the center of mass condition $-\sum_{\ell \neq k} m_\ell q_\ell = m_k q_k$, that

$$\sum_{\ell \neq k} m_\ell \left(\frac{1}{\|q_k - q_\ell\|^3} - \frac{c}{m_1 + m_2 + m_3} \right) (q_k - q_\ell) = 0 \qquad (k = 1, 2, 3).$$

By hypothesis, the two vectors $q_k - q_\ell$ in this sum are linearly independent, so their coefficients $\|q_k - q_\ell\|^{-3} - c/(m_1 + m_2 + m_3)$ have to vanish. Therefore all sides $\|q_k - q_\ell\|$ are of equal length. □

11.32 Remarks (Central Configurations)

1. It is surprising that even in the case of unequal masses, one obtains an equilateral triangle.
2. Concerning these periodic solutions to the 3-body problem found by himself, Lagrange wrote in [Lag], page 230: “Cette recherche n’est à la vérité que de pure curiosité”.[14] As a matter of fact, it was only in 1906 that celestial bodies were found that orbit near a Lagrange point. These objects, called Trojans, are populating a neighborhood of the Lagrange points L_4 and L_5 for the pair of sun and its heaviest planet Jupiter. Today, it is believed that there are a million Trojans of diameter larger than 1 kilometre.
   For the pair consisting of earth and sun, the (unstable) Lagrange points L_1 and L_2 are used as positions for satellites.
3. For $n > 3$ particles, the structure of central configurations gets complicated rather quickly. For $n = 4$ mass points, it has been known since the article [HaMo] by HAMPTON and MOECKEL that there are only finitely many equivalence classes of planar central configurations (probably between 32 and 50, depending on the relative masses). In [AlK], ALBOUY and KALOSHIN prove similar results for $n = 5$.
4. The five Lagrange points lead to bifurcations of the mapping defined by the constants of motion, and to a change in topology for its pre-images.
   STEVEN SMALE in [Sm2] has studied the n-body problem with regard to such bifurcations. ◊

The Lagrange points serve as a starting point for a better understanding of the 3-body problem. This can probably be seen most easily in connection with the *restricted 3-body problem*, in which one mass moves within the gravitational field of two stars that move on Kepler ellipses about their common center of mass. For the circular case, this problem therefore differs from the integrable two center problem by a rotation with constant angular velocity (Figure 11.3.3).

A better understanding of the non-restricted planar 3-body problem can be obtained by considering the space of its equivalent configurations. In doing so, we classify the configurations $q = (q_1, q_2, q_3) \in \widehat{M}_0$ into equivalence classes of triangles under rotations and dilations. The manifold obtained in this manner has the topology of a sphere with three punctures, $S^2 \setminus \{p_1, p_2, p_3\}$, called *form sphere*, see Figure 11.3.4.

This can be seen by interpreting the difference vectors $q_2 - q_1$ and $q_3 - q_1$ as nonzero complex numbers. The ratio $\frac{q_3-q_1}{q_2-q_1}$ characterizes the equivalence class. All points in $\mathbb{C}\setminus\{0, 1\}$ occur as ratios. By stereographic projection, $\mathbb{C}\setminus\{0, 1\}$ corresponds to the triply punctured sphere. These three puncture points correspond to the three 2-body collisions.

[14] Translation: This study is actually made just for the sake of mere curiosity.

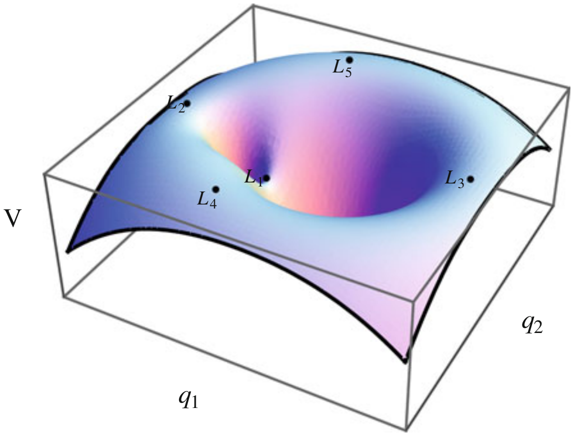


Figure 11.3.3 Potential of the restricted 3-body problem, with the five Lagrange points

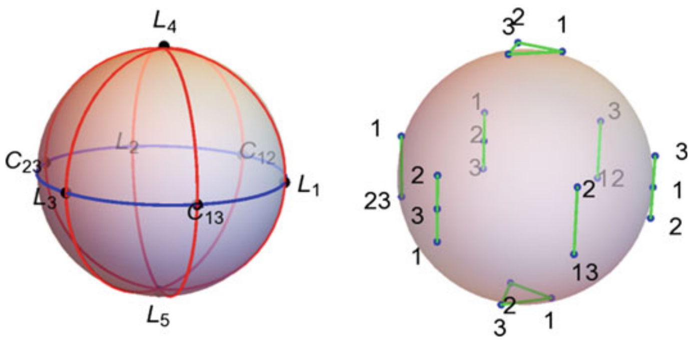


Figure 11.3.4 The form sphere, with the five Lagrange points $L_1, \ldots, L_5$ and the three collision points C_{kl}. The equator parametrizes the collinear configurations, the other three great circles parametrize isosceles triangles

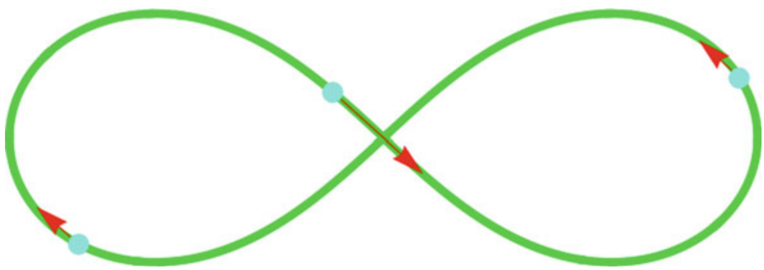


Figure 11.3.5 A stable solution of the 3-body problem, after CHENCINER and MONTGOMERY [CM, Mon2]

11.33 Definition *A planar T-periodic solution*

$$t \mapsto q(t) = \big(q_1(t), \dots, q_n(t)\big) \in \widehat{M}_0$$

of the n-body problem is called a **choreography** *if the orbits of the n mass points satisfy*

$$q_{k+1}(t) = q_1(t - Tk/n) \qquad (t \in \mathbb{R},\ k = 1, \dots, n-1),$$

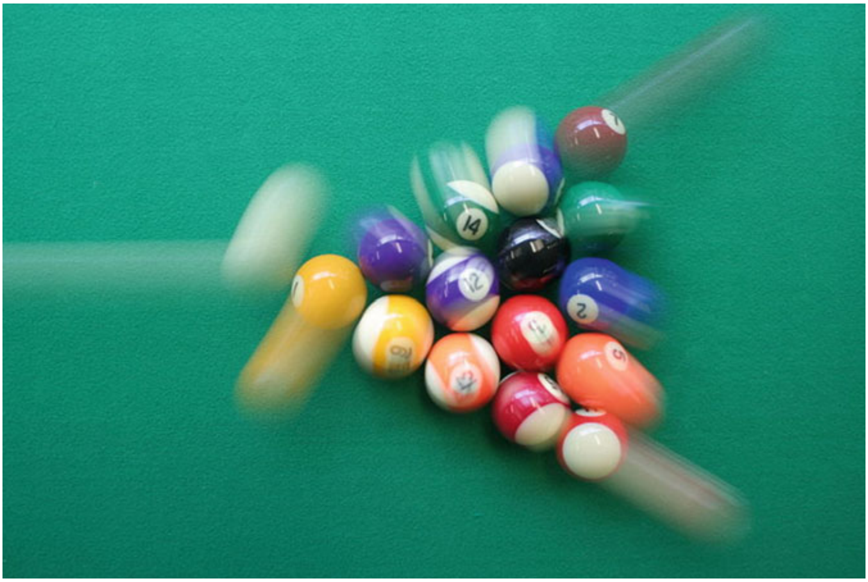
i.e., if all points traverse the same orbit in the plane, with a fixed distance in time.

In the past years, numerous choreographies have been found, in particular one in which three equal masses move through a figure 8 shape (see Figure 11.3.5). This solution is stable, so it is conceivable that somewhere in the universe three stars actually do perform this curious dance.

More importantly, this and other choreographies may possibly open a geometric access to the 3-body problem: Viewed as a curve on the form sphere, this Figure 8 is wrapped around the three collision singularities along the equator. The general question is which homotopy classes of collision free closed curves on the form sphere have representatives that are actually solutions.

11.34 Literature The literature on the n-body problem of celestial mechanics is vast. The textbooks [AM, DH, HaMe, Mos4, SM] offer introductions to its various aspects; [Sa] is distinguished by a unique selection of subjects. $\diamondsuit$

Chapter 12
Scattering Theory



Billiard[1]

The major part of our knowledge about molecules, atoms, and elementary particles is obtained by scattering experiments, in which particles of a specific initial velocity collide with each other or with a fixed target. After the scattering process, one registers which particles occur and with which velocity. Although the correct language for the description of these processes is quantum mechanics, in certain situations, its predictions agree with those of classical mechanics in a good approximation.

[1]Image: https://commons.wikimedia.org/wiki/File:Billard.JPG, August 2006, photo by Noé Lecocq in collaboration with H. Caps. Courtesy of Noé Lecocq.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_12

Examples of genuinely classical scattering processes are given by billiard balls, or by comets in the gravitational field of our solar system.

12.1 Scattering in a Potential

"Therefore rectilinear motion occurs only to things that are not in proper condition and are not in complete accord with their nature, when they are separated from their whole and forsake its unity."[2] NICOLAUS COPERNICUS, in *De Revolutionibus Orbium Coelestium* (1543) [Cop]

We first study the scattering of a (classical) particle in a long range potential. This process is described by the Hamiltonian

$$H(p,q) := \tfrac{1}{2}\|p\|^2 + V(q) \quad \text{on the phase space } P := \mathbb{R}^d_p \times \mathbb{R}^d_q\,, \tag{12.1.1}$$

where the potential $V \in C^2(\mathbb{R}^d_q, \mathbb{R})$ tends to 0 at infinity. In scattering theory, we distinguish short and long range potentials. We can define them in terms of the smoothed absolute value function

$$\langle \cdot \rangle : \mathbb{R}^d \to [1,\infty) \quad , \quad \langle q \rangle := \sqrt{\|q\|^2 + 1}$$

and multi-indices $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$, $|\alpha| := \alpha_1 + \ldots + \alpha_d$ and partial derivatives $\partial^\alpha := \frac{\partial^{\alpha_1}}{\partial^{\alpha_1}_{q_1}} \ldots \frac{\partial^{\alpha_d}}{\partial^{\alpha_d}_{q_d}}$ indexed by multi-indices as follows:

12.1 Definition (Classes of Potentials) $V \in C^2(\mathbb{R}^d, \mathbb{R})$ is called

- **long range** if it satisfies, for appropriate $\varepsilon > 0$, $c > 0$, the estimate

$$|\partial^\alpha V(q)| \le c\,\langle q \rangle^{-|\alpha|-\varepsilon} \qquad (q \in \mathbb{R}^d, |\alpha| \le 2)\,;$$

- **short range** if it satisfies, for appropriate $\varepsilon > 0$, $c > 0$, the estimate

$$|\partial^\alpha V(q)| \le c\,\langle q \rangle^{-|\alpha|-1-\varepsilon} \qquad (q \in \mathbb{R}^d, |\alpha| \le 2).$$

12.2 Remarks (Range and Asymptotics)

1. So if V is short range, then it is in particular also long range, with the same constants ε and c.
2. As shown in Theorem 12.11 below, short range potentials have the property that their scattering trajectories are asymptotic to the orbit curves

$$\Phi_t^{(0)}(x_0) = \big(p^{(0)}(t,x_0), q^{(0)}(t,x_0)\big) = \big(p_0, q_0 + p_0\, t\big) \qquad (t \in \mathbb{R}) \tag{12.1.2}$$

[2] Quoted after Nicholas Copernicus: On the Revolutions, edited by Jerzy Dobrzycki, translation and commentary by Edward Rosen, 1978, The Johns Hopkins University Press; p 17.

of *free motion* as defined by

$$H^{(0)} : P \to \mathbb{R} \quad , \quad H^{(0)}(p, q) := \tfrac{1}{2}\|p\|^2$$

with initial value $x_0 = (p_0, q_0) \in P$.

Certain long range potentials do not have this property, namely the *Coulomb* potential $q \mapsto -Z/\|q\|$, which is important in physics and is called *Kepler* potential when $Z > 0$. Whereas the Kepler hyperbola is asymptotic to an unparametrized straight line, it is traversed with a speed $\sqrt{2(E + Z/\|q\|)}$, which deviates too much from the free speed $\sqrt{2E}$.

3. Long range potentials give rise to a force $F(q) := -\nabla V(q)$, that is of order $F(q) = \mathcal{O}\big(\langle q\rangle^{-1-\varepsilon}\big)$, and is therefore radially integrable. As we shall see, this guarantees at least the asymptotics of the velocity. ♢

Now assume that the potential V is long range. From the assumed continuity and $\lim_{\|q\|\to\infty} V(q) = 0$, it follows that the infimum $V_{\min} := \inf_{q\in\mathbb{R}^d_q} V(q) > -\infty$ is finite, and we obtain a complete flow for (12.1.1) by Theorem 11.1:

$$\Phi \in C^1(\mathbb{R} \times P, P) \quad , \quad \big(p(t,x), q(t,x)\big) := \Phi_t(x) := \Phi(t,x)\,. \qquad (12.1.3)$$

Now we want to obtain more precise information about this dynamics, and we begin by distinguishing scattering and bound orbits.

12.3 Definition *We distinguish the following kinds of subsets of phase space:*
Bound states $b := b^+ \cap b^-$ with $b^\pm := \Big\{x \in P \;\Big|\; \limsup_{t\to\pm\infty} \|q(t,x)\| < \infty\Big\}$,

Scattering states $s := s^+ \cap s^-$ with $s^\pm := \Big\{x \in P \;\Big|\; \lim_{t\to\pm\infty} \|q(t,x)\| = \infty\Big\}$
and
Trapped states $t := t^+ \cup t^-$ with $t^\pm := b^\pm \cap s^\mp$.
For energy surfaces $\Sigma_E = H^{-1}(E)$ with $E \in \mathbb{R}$, we set $b^\pm_E := b^\pm \cap \Sigma_E$ etc.

12.4 Remarks

1. The sets defined here are Φ-invariant.
2. For energy $E < 0$ the energy surface Σ_E is compact since $\lim_{\|q\|\to\infty} V(q) = 0$ and $H(p,q) \ge V(q)$. So we have $\Sigma_E = b_E$ for those energies.
   In the sequel, we omit the discussion of energy $E = 0$ and turn to energies $E > 0$. ♢

12.5 Theorem (Scattering Asymptotics) *Let V be a long range potential.*

1. *Then for appropriate* $c_1 > 0$, *the* **virial radius** $R_{\text{vir}}(E) := c_1 E^{-1/\varepsilon}$ $(E > 0)$ has the property that $x_0 = (p_0, q_0) \in \Sigma_E$ with $\langle q_0\rangle \ge R_{\text{vir}}(E)$ and $\langle q_0, p_0\rangle \ge 0$ implies $x_0 \in s^+$.

2. *For all initial conditions* $x_0 \in s_E^{\pm}$, *the* **asymptotic momenta**

$$p^{\pm}(x_0) := \lim_{t\to\pm\infty} p(t, x_0) \in \mathbb{R}^d$$

exist and satisfy $\|p^{\pm}(x_0)\| = \sqrt{2E}$. *We call* $\hat{p}^{\pm}(x_0) := \frac{p^{\pm}(x_0)}{\|p^{\pm}(x_0)\|} \in S^{d-1}$ *the* **asymptotic direction**.

3. *If* V *is short range, then the* **asymptotic impact parameters**

$$q_{\perp}^{\pm} : s_E^{\pm} \to \mathbb{R}^d \quad , \quad q_{\perp}^{\pm}(x_0) = \lim_{t\to\pm\infty} \Big(q(t, x_0) - \big\langle q(t, x_0), \hat{p}^{\pm}(x_0) \big\rangle \, \hat{p}^{\pm}(x_0) \Big)$$

exist.

Proof:
• Without loss of generality, we only consider the limit $t \to +\infty$ and write the solution for initial value x_0 in the form $t \mapsto \big(p(t), q(t)\big)$. For

$$F(t) := \tfrac{1}{2}\|q(t)\|^2 , \quad \text{we then have } F'(t) = \langle q(t), p(t) \rangle$$

and, using the constant energy $E = H(x_0) = H\big(p(t), q(t)\big)$,

$$F''(t) = \|p(t)\|^2 - \langle q(t), \nabla V(q(t)) \rangle = 2E - 2V\big(q(t)\big) - \big\langle q(t), \nabla V\big(q(t)\big)\big\rangle . \tag{12.1.4}$$

With the long range property of V, this implies

$$F''(t) \ge 2E - 2\big|V\big(q(t)\big)\big| - \langle q(t) \rangle \; \big\|\nabla V\big(q(t)\big)\big\| \ge 2E - c(2+d)\,\langle q(t) \rangle^{-\varepsilon} , \tag{12.1.5}$$

with the constants ε and c from Definition 12.1. Now if $\langle q(s) \rangle \ge R_{\text{vir}}(E)$ with

$$R_{\text{vir}}(E) = c_1\, E^{-1/\varepsilon} \quad \text{and} \quad c_1 := \big(c\,(2+d)\big)^{1/\varepsilon}, \tag{12.1.6}$$

then the radial acceleration satisfies $F''(s) \ge E > 0$. This implies $F'(t) \ge F'(0) + Et$, as long as the condition $\langle q(s) \rangle \ge R_{\text{vir}}(E)$ is satisfied for all $s \in [0, t]$. As $F'(0) = \langle q_0, p_0 \rangle \ge 0$ and $\langle q(0) \rangle \ge R_{\text{vir}}(E)$, it follows under this hypothesis that

$$F'(s) \ge Es \qquad \big(s \in [0, t]\big),$$

hence

$$F(t) \ge F(0) + \int_0^t F'(s)\, \mathrm{d}s \ge \tfrac{1}{2}\Big[\big(R_{\text{vir}}(E)\big)^2 + Et^2\Big]. \tag{12.1.7}$$

Now let $T := \inf\{s > 0 \mid F'(s) = 0\}$. From the Taylor expansion $F'(s) = F'(0) + F''(0)s + o(s) \ge Es + o(s)$, we infer $T > 0$. If we had $T \in (0, \infty)$, then it would follow that $\langle q(T) \rangle = \langle q(0) \rangle + \int_0^T \frac{F'(s)}{\langle q(s) \rangle}\, \mathrm{d}s \ge R_{\text{vir}}(E)$ and $F'(T) =$

$F'(0) + \int_0^T F''(s)\,\mathrm{d}s \geq E\,T > 0$, in contradiction to the definition of T.

Therefore it follows that $\lim_{t\to\infty} \|q(t)\| = \infty$, i.e., $x_0 \in s^+$.

• For every $x_0 = (p_0, q_0) \in s_E^+$, there exists a time t_0 with $\langle q(t_0), p(t_0)\rangle \geq 0$ and $\langle q(t_0)\rangle \geq R_{\mathrm{vir}}(E)$. We will assume without loss of generality that $t_0 = 0$. By (12.1.7), we then have $\langle q(t)\rangle \geq \sqrt{(R_{\mathrm{vir}}(E))^2 + Et^2}$ for all $t \geq 0$, and thus for all $t_2 \geq t_1 > 0$,

$$\|p(t_2) - p(t_1)\| = \left\| \int_{t_1}^{t_2} \nabla V\big(q(s)\big)\,\mathrm{d}s \right\| \leq dc \int_{t_1}^{t_2} \langle q(s)\rangle^{-1-\varepsilon}\,\mathrm{d}s \qquad (12.1.8)$$

$$\leq d\,c \int_{t_1}^{\infty} \left[\big(R_{\mathrm{vir}}(E)\big)^2 + Es^2\right]^{-\frac{1+\varepsilon}{2}}\,\mathrm{d}s$$

$$\leq d\,c\,E^{-\frac{1+\varepsilon}{2}} \int_{t_1}^{\infty} s^{-(1+\varepsilon)}\,\mathrm{d}s = c_2\, t_1^{-\varepsilon} \text{ with } c_2 := \frac{d\,c\,E^{-\frac{1+\varepsilon}{2}}}{\varepsilon}.$$

Thus the Cauchy condition for the existence of $p^+(x_0) = \lim_{t\to+\infty} p(t)$ is satisfied.

• Because $\big(p(t), q(t)\big) \in \Sigma_E$, hence $\|p(t)\| = \sqrt{2\big(E - V(q(t))\big)}$, and because $\lim_{t\to+\infty} V\big(q(t)\big) = 0$, it also follows that

$$\|p^+(x_0)\| = \lim_{t\to+\infty} \|p(t)\| = \sqrt{2E} > 0\,.$$

• For $q_\perp(t) := q(t) - \big\langle q(t), \hat{p}^+\big\rangle\,\hat{p}^+$ (with $\hat{p}^+ := \hat{p}^+(x_0)$), one has

$$q_\perp(t_2) - q_\perp(t_1) = \int_{t_1}^{t_2} \left[p(s) - \big\langle p(s), \hat{p}^+\big\rangle\,\hat{p}^+\right]\mathrm{d}s$$

$$= \int_{t_1}^{t_2} \left[(p(s) - p^+) - \big\langle p(s) - p^+, \hat{p}^+\big\rangle\,\hat{p}^+\right]\mathrm{d}s\,,$$

and thus for $t_2 \geq t_1 > 0$ in the short range case, analogous to (12.1.8):

$$\|q_\perp(t_2) - q_\perp(t_1)\| \;\leq\; \int_{t_1}^{\infty} \|p(s) - p^+\|\,\mathrm{d}s \;\leq\; c_3 \int_{t_1}^{\infty} s^{-1-\varepsilon}\,\mathrm{d}s \;=\; \frac{c_3}{\varepsilon}\, t_1^{-\varepsilon}$$

with $c_3 := \frac{d\,c\,E^{-\left(1+\frac{\varepsilon}{2}\right)}}{1+\varepsilon}$, because

$$\|p(t) - p^+\| \;\leq\; d\,c \int_t^{\infty} \langle q(s)\rangle^{-2-\varepsilon}\,\mathrm{d}s \;\leq\; d\,c \int_t^{\infty} (Es^2)^{-(1+\frac{\varepsilon}{2})}\,\mathrm{d}s \;=\; c_3\, t^{-(1+\varepsilon)}.$$

So the asymptotic impact parameter $q_\perp^+(x_0) = \lim_{t\to+\infty} q_\perp(t)$ also satisfies the Cauchy condition. □

12.6 Remarks (Asymptotic Momenta and Angular Momenta)

1. Similar to the asymptotic velocities in a periodic potential (see Theorem 11.7), the asymptotic momenta can be written as Cesàro limits. So if $p^{\pm}(x_0) = \lim_{t\to\pm\infty} p(t, x_0)$ converges, one has

$$p^{\pm}(x_0) = \lim_{T\to+\infty} \frac{1}{T}\int_0^T p(\pm t, x_0)\,\mathrm{d}t = \lim_{t\to\pm\infty} \frac{q(t, x_0)}{t} \qquad (12.1.9)$$

The latter limit exists for *all* initial values $x_0 \in P$. In contradistinction to Theorem 11.7 however, we here have typically $p^+(x_0) \neq p^-(x_0)$ (provided $d \geq 2$ and the energy $H(x_0)$ is positive).

2. For short range potentials, Theorem 12.5 implies that the limit

$$L^{\pm}(x_0) := \lim_{t\to\pm\infty} q(t, x_0) \wedge p(t, x_0) \qquad \left(x_0 \in s_E^{\pm}\right) \qquad (12.1.10)$$

exists; it describes the *asymptotic angular momenta*. In dimension $d = 2$, the 2-form $q \wedge p$ can be identified with the number $q_1 p_2 - q_2 p_1$. In dimension $d = 3$, it can be identified with the vector $q \times p$. ♢

12.7 Exercises (Asymptotics of Scattering in a Potential)

1. Show that the Cesàro limits in (12.1.9) exist for *all* initial conditions $x_0 \in P$ (whereas for $x_0 \in b^{\pm}$, the limits $\lim_{t\to\pm\infty} p(t, x_0)$ do not exist in general, and for the motion in periodic potentials, even the Cesàro limits do not always exist). This property is called *asymptotic completeness*, see Theorem 12.53.
2. Show that the asymptotic angular momenta (12.1.10), and with them also the asymptotic impact parameters, also exist if V differs by a short range potential from some long range, but centrally symmetric, potential W.

An analogous claim holds for the singular n-atom molecule potentials

$$V(q) := -\sum_{k=1}^{n} \frac{Z_k}{\|q - s_k\|} \quad \text{and} \quad W(q) := -\frac{Z}{\|q\|} \qquad (12.1.11)$$

with atomic sites $s_k \in \mathbb{R}^3$ and total charge $Z := \sum_{k=1}^{n} Z_k$. ♢

12.8 Corollary *Let V be a long range potential.*
• *For all $E > 0$, the energy shell is the disjoint union*

$$\Sigma_E = b_E \,\dot\cup\, s_E \,\dot\cup\, t_E \qquad (12.1.12)$$

of bound, scattering, and trapped states. Here b_E is compact and s_E is open.

• *There exists an energy threshold* $E_0 > 0$ *such that for all* $E \geq E_0$, *there only exist scattering states, i.e.,*

$$\Sigma_E = s_E \,. \tag{12.1.13}$$

Proof:
• We first show that $\Sigma_E = s_E^+ \dot{\cup} b_E^+$. It is immediate from their definition that the sets s_E^+ and b_E^+ are disjoint.

• If some $x_0 \in \Sigma_E$ is not contained in b^+, then there exists an increasing sequence of times $(t_n)_{n\in\mathbb{N}}$ for which $\|q(t_{n+1})\| \geq \|q(t_n)\|$ and $\lim_{n\to\infty} \|q(t_n)\| = \infty$. So there exists n with $\|q(t_n)\| \geq R_{\text{vir}}(E)$ and, since $F(t_{n+1}) > F(t_n)$ for $F(t) := \frac{1}{2}\|q(t)\|^2$, there exists a time $t \in [t_n, t_{n+1}]$ with $F'(t) = \langle q(t), p(t)\rangle \geq 0$ and $\|q(t)\| \geq R_{\text{vir}}(E)$. Concerning the phase space point $x(t) = (p(t), q(t)) \in \Sigma_E$, Theorem 12.5 tells us that $x(t) \in s^+$. As the set s^+ is invariant under Φ, the initial point x_0 is in s^+ as well.

• The analogous decomposition of the energy surface $\Sigma_E = s_E^- \dot{\cup} b_E^-$ determined by the dynamics in the past follows from Theorem 11.5 about reversible flows.

• Combining both decompositions, we obtain

$$\Sigma_E = (s_E^+ \cap s_E^-) \dot{\cup} (s_E^+ \cap b_E^-) \dot{\cup} (b_E^+ \cap s_E^-) \dot{\cup} (b_E^+ \cap b_E^-) = s_E \dot{\cup} t_E^- \dot{\cup} t_E^+ \dot{\cup} b_E \,,$$

hence by $t_E = t_E^- \dot{\cup} t_E^+$ claim (12.1.12).
• We let $E_0 := c_1^\varepsilon$ with the constant c_1 from (12.1.6), so that $R_{\text{vir}}(E_0) = 1$, and we consider the time dependence of $F(t) = \frac{1}{2}\|q(t)\|^2$ for $E \geq E_0$ and initial value $x_0 = (p_0, q_0) \in \Sigma_E$.

As now the inequality $F''(t) \geq 2E - E_0 \langle q(t)\rangle^{-\varepsilon} \geq E > 0$ (see (12.1.5)) holds for all $x_0 \in \Sigma_E$, it follows that

$$F(t) \geq F(0) + F'(0)t + \frac{E}{2}t^2,$$

hence $\lim_{t\to\pm\infty} \|q(t)\| = \infty$, i.e., $x_0 \in s_E$, and therefore (12.1.13).
• b_E is contained in the bounded set $\{(p, q) \in \Sigma_E \mid \|q\| \leq R_{\text{vir}}(E)\}$. If x_∞ is the limit of a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in b_E, then continuity of the flow implies that

$$\Phi_t(x_\infty) = \lim_{n\to\infty} \Phi_t(x_n) \qquad (t \in \mathbb{R}). \tag{12.1.14}$$

Now if we had $x_\infty \in s_E^+$, then for an appropriate time t, it would follow that

$$\langle q(t, x_\infty)\rangle > R_{\text{vir}}(E) \quad , \quad \langle q(t, x_\infty), p(t, x_\infty)\rangle > 0\,.$$

By (12.1.14), the same would apply for $\Phi_t(x_n)$ with n large. But then, x_n would also be contained in s_E^+. Therefore we must have $x_\infty \in b_E^+$, and similarly $x_\infty \in b_E^-$; so b_E is closed and therefore compact.
• Conversely, this implies that the $s_E^\pm$ are open, hence also $s_E = s_E^+ \cap s_E^-$. □

So this corollary rules out the existence of orbits for which

$$\limsup_{t\to+\infty} \|q(t)\| = \infty \quad \text{, but} \quad \liminf_{t\to+\infty} \|q(t)\| < \infty\,, \tag{12.1.15}$$

and likewise the analogous statement for $t \to -\infty$.

We summarize the informations obtained about the asymptotics of scattering orbits in Theorem 12.5 as follows: Denoting by

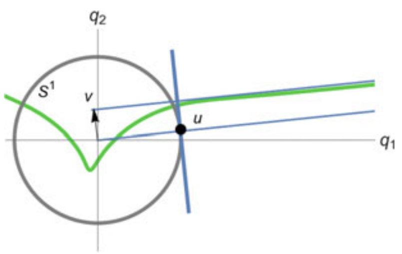
$$TS^{d-1} := \{(u,v) \in \mathbb{R}^d \times \mathbb{R}^d \mid \|u\| = 1, \langle u, v\rangle = 0\} \tag{12.1.16}$$

the *tangent bundle* of the $(d-1)$-sphere S^{d-1} (see page 499), we have found mappings

$$s_E^{\pm} \to TS^{d-1} \quad , \quad x \mapsto \big(\hat{p}^{\pm}(x), q_{\perp}^{\pm}(x)\big)\,.$$

The manifold TS^{d-1} parametrizes the oriented lines in $\mathbb{R}^d$, because such lines can uniquely be represented in the form

$$t \mapsto v + ut \quad \text{with} \quad (u,v) \in TS^{d-1}\,.$$



The trajectory $t \mapsto q(t)$ in configuration space is asymptotic to the lines $s \mapsto q_{\perp}^{\pm} + s\,p^{\pm}$ in the sense that in the limit $t \to \pm\infty$, the minimal distances

$$d^{\pm}\big(q(t)\big) := \min_{s\in\mathbb{R}} \Big\| q(t) - (q_{\perp}^{\pm} + s\,p^{\pm}) \Big\|$$

tend to zero. This is because

$$d^{\pm}\big(q(t)\big) = \|q_{\perp}^{\pm}(t) - q_{\perp}^{\pm}\|\,,$$

where the quantity $q_{\perp}^{\pm}(t) := q(t) - \big\langle q(t), \hat{p}^{\pm}\big\rangle\,\hat{p}^{\pm}$ is taken from Theorem 12.5.

12.9 Remark (Long Range Potentials)
Such an asymptotics of the scattering trajectory towards straight lines can also occur in long range potentials, see Exercise 12.7.2. Just think of the Kepler hyperbolas.

However, according to Remark 12.2, these Kepler hyperbolas are traversed at speed $t \mapsto \sqrt{2\left(E + \frac{Z}{\|q(t)\|}\right)}$. This differs from the speed $\sqrt{2E}$ for the free motion generated by $H^{(0)}(p,q) = \frac{1}{2}\|p\|^2$ so much that an *asymptotic synchronization* becomes impossible, i.e., there does not exist t_0 for which

$$\lim_{t\to+\infty} \|q(t) - (q_{\perp}^{+} + (t-t_0)p^{+})\| = 0\,,$$

see Exercise 12.10.1. $\Diamond$

12.10 Exercises (Scattering Orbits)

1. Show that for the Kepler hyperbolas of energy $E > 0$ with angular momentum $\ell \in \mathbb{R}$ in the potential $q \mapsto -Z/\|q\|$, the time elapsing between two radii is given by the corresponding definite integral for the indefinite integral

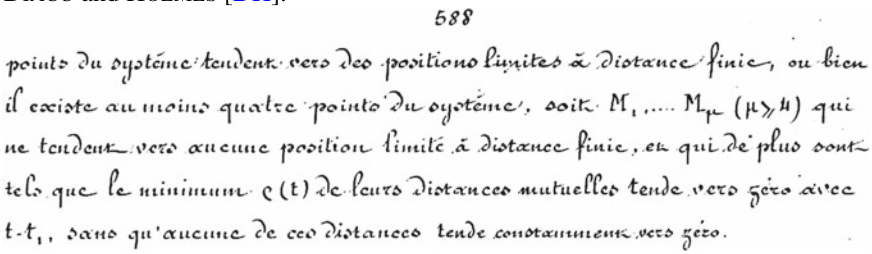
$$\int \frac{r}{\sqrt{2r^2E + 2Zr - \ell^2}}\,\mathrm{d}r = \frac{r}{\sqrt{2E}}\sqrt{1 + \frac{Z}{rE} - \frac{\ell^2}{2r^2E}} - \frac{Z}{(2E)^{3/2}} \ln\left(Er + \tfrac{1}{2}Z + \sqrt{E(r^2E + Zr - \tfrac{1}{2}\ell^2)}\right) + c\,. \tag{12.1.17}$$

The logarithmic term in (12.1.17) makes asymptotic synchronization impossible (see also THIRRING [Th1], Chapter 4.2).

Painlevés Stockholm Lectures

In 1895, upon invitation by King Oscar II of Sweden, the 31 year old mathematician (and subsequent prime minister of France) PAUL PAINLEVÉ gave a lecture series on differential equations in Stockholm.

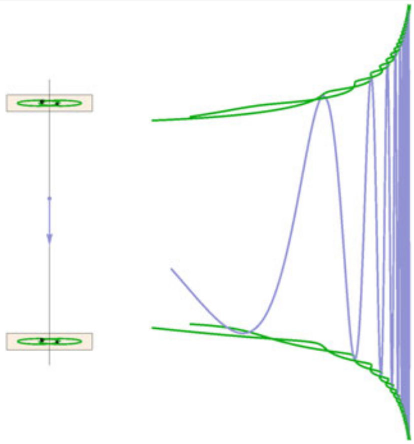
As the highlight and conclusion of his manuscript [Pai] of nearly 600 pages, which was published two years later, he pointed out that in celestial mechanics, some kinds of singularities other than collisions could be conceivable. See the manuscript,[3] and DIACU and HOLMES [DH].


588

points du systéme tendent vers des positions limites à distance finie, ou bien il existe au moins quatre points du systéme, soit $M_1, \ldots\ldots M_\mu$ $(\mu \geq 4)$ qui ne tendent vers aucune position limite à distance finie, et qui de plus sont tels que le minimum $\rho(t)$ de leurs distances mutuelles tende vers zéro avec $t - t_1$, sans qu'aucune de ces distances tende constamment vers zéro.

[3] '…points du systéme tendent vers des positions limites à distance finie, ou bien il existe au moins quatre points du systéme, soit $M_1, \ldots, M_\mu$ $(\mu \geq 4)$ qui ne tendent vers aucune position limite à distance finie, et qui de plus sont tels que le minimum $\rho(t)$ de leurs distances mutuelles tende vers zéro avec $t - t_1$, sans qu'aucune de ces distances tende constamment vers zéro.'
Translation: 'mass points tend to positions with a finite distance, or else there exist at least four mass points $M_1, \ldots, M_\mu$, $(\mu \geq 4)$, that do not converge to a limit position within finite distance, and for which their mutual minimum distance $\rho(t)$ goes to 0 as $t - t_1$ goes to 0, but without any single one of these distances converging to 0.'
[In Xia's example, some distances converge, but not all do.].

This conjecture was proved by JEFF XIA in [Xi] almost 100 years later. He showed in 1992 that in the n-body problem (1.8) of celestial mechanics, mass points can escape to infinity in finite time. In the configuration he found, a 'messenger' star shuttles between two double star systems increasingly more rapidly and drives them to infinity in finite time. This obviously is not an example of astronomical reality, but it shows what kind of surprises need to be expected when studying the n-body problem mathematically.



Xia's solution for the 5-body problem of celestial mechanics

There also exist analogous solutions that have the divergent behavior only in the limit of infinite time. In this case, the position of the messenger star satisfies (12.1.15), and the velocities do not have a limit, in other words, the motion is not asymptotically complete in the sense of Definition 12.40.
Sometimes in mathematics, one has to wait a while for a proof.

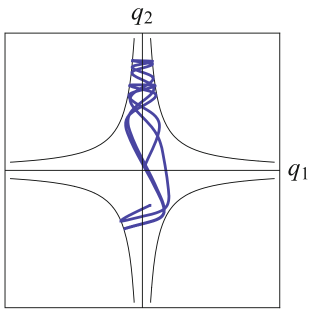
2. For $E > 0$, we consider the energy surface $\Sigma_E \subseteq P := \mathbb{R}^2_p \times \mathbb{R}^2_q$ of the Hamiltonian[4] $H : P \to \mathbb{R}\,,\ H(p,q) = \frac{1}{2}\|p\|^2 + V(q)$ with the potential

$$V(q) := \tfrac{1}{2} q_1^2 q_2^2\,.$$

Show:
Except for the four orbits in the closed set

$$s_E := \big\{(p,q) \in \Sigma_E \mid q_1 q_2 = p_1 p_2 = 0,\ q \parallel p\big\},$$

there are no initial conditions $x_0 \in \Sigma_E$ for which $\lim_{t\to+\infty} \|q(t,x_0)\| = \infty$. In particular, the set s_E of scattering orbits is not open.



A typical trajectory in Hill's region $V(q) \leq E$

Hint:
Assuming $\lim_{t\to+\infty} q_1(t,x_0) = +\infty$, consider the quantity $J(p,q) := \frac{1}{2}\left(q_1 q_2^2 + \frac{p_2^2}{q_1}\right)$ and show that it is approximately constant (see Section 15.2). Conclude by comparison with H that this quantity is actually 0. $\Diamond$

[4]This Hamiltonian occurs in the proof of non-integrability for the classical Yang-Mills equation with gauge group SU(2).

12.2 The Møller Transformations

We now compare the flow (12.1.3) generated by H from (12.1.1) with the free dynamics (12.1.2) generated by $H^{(0)}$.

12.11 Theorem

• *For short range potentials* $V \in C^2\big(\mathbb{R}^d, \mathbb{R}\big)$, *the* **Møller transformations**[5] *(compare with Remark 12.2.2 and Figure 12.2.1)*

$$\Omega^\pm : P_+^{(0)} \to s^\pm \quad , \quad \Omega^\pm := \lim_{t\to\pm\infty} \Phi_{-t} \circ \Phi_t^{(0)} \tag{12.2.1}$$

are homeomorphisms on the phase space domain $P_+^{(0)} := \big(\mathbb{R}_p^d \setminus \{0\}\big) \times \mathbb{R}_q^d$, *and they conserve energy:*

$$H \circ \Omega^\pm(x) = H^{(0)}(x) \qquad \big(x \in P_+^{(0)}\big). \tag{12.2.2}$$

• *With respect to time reversal* $\mathcal{T}$ *(Definition 11.3), one has* $\Omega^- = \mathcal{T} \circ \Omega^+ \circ \mathcal{T}$.
• *The Møller transformations conjugate the two flows:*

$$\Omega^\pm \circ \Phi_t^{(0)} = \Phi_t \circ \Omega^\pm \qquad (t \in \mathbb{R}). \tag{12.2.3}$$

• *If the potential is even smooth,* $V \in C^\infty\big(\mathbb{R}^d, \mathbb{R}\big)$, *and has compact support, then* $\Omega^\pm$ *are smooth, symplectic and therefore volume preserving mappings.*

12.12 Remarks (Variants)

1. Working harder than is done in the proof below, one can even prove the differentiability of the Møller transformations for short range potentials under the hypotheses of Theorem 12.11.
2. If we were to strengthen Definition 12.1 of short range potentials V to the effect that for $V \in C^\infty(\mathbb{R}^d, \mathbb{R})$ we demand with appropriate $c_\alpha > 0$ that

$$|\partial^\alpha V(q)| \le c_\alpha \langle q\rangle^{-|\alpha|-1-\varepsilon} \qquad \big(q \in \mathbb{R}^d,\ \alpha \in \mathbb{N}_0^d\big)\,,$$

   then the Møller transformations would also be C^∞-diffeomorphisms.
3. If in contrast, one were to weaken Definition 12.1 to the effect of omitting the decay condition for the *second* derivative, one can find potentials for which different orbits have the same asymptotic data (see Simon [Sim]). This phenomenon is comparable to the existence of several solutions for the initial value problem in the case of differential equations that are not Lipschitz continuous, see Example 3.11.2. ◊

[5] *Christian Møller*, Danish physicist (1904–1980). See his article:
General properties of the characteristic matrix in the theory of elementary particles. Danske Vid. Selsk. Mat.-Fys. Medd. **23**, (1945) 1–48.

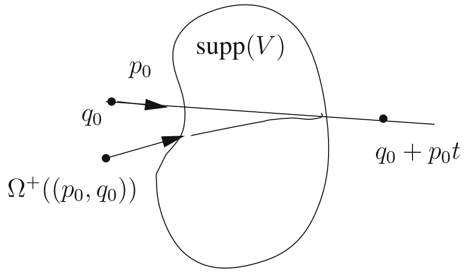


Figure 12.2.1 Definition of the Møller transformation Ω^+ for a potential V with compact support

Proof of Theorem 12.11:

• We begin with the last claim, which is easiest to prove. Outside the phase space area over the compact support $\text{supp}(V) = \overline{\{q \in \mathbb{R}^d \mid V(q) \neq 0\}}$ of V, both flows coincide. Therefore for all $x_0 = (p_0, q_0) \in P_+^{(0)}$, there is a time $T \geq 0$ so that for all initial conditions $x = (p, q)$ in $U(x_0) := \big\{x \in P_+^{(0)} \mid \|x - x_0\| < 1/2\, \|p_0\|\big\}$, one has

$$q + tp \notin \text{supp}(V) \qquad (t \geq T).$$

So for $x \in U(x_0)$, the limit in (12.2.1) is reached already at time T. As both flows are symplectic, the assertion is proved.

For all $x = (p, q) \in U(x_0)$, the velocity satisfies $\|p\| > \frac{1}{2}\|p_0\| > 0$. On the other hand, $\text{supp}(V)$ lies in the interior of some ball $B_R^d = \{q \in \mathbb{R}^d \mid \|q\| \leq R\}$ of a large radius $R > 0$, whereas the distance of the initial point q to the center 0 of the ball is less than $\|q_0\| + \frac{1}{2}\|p_0\|$. Therefore the free flow will have led the set $U(x_0)$ out of the phase space area above $\text{supp}(V)$ forever after a time of at least

$$T \equiv T(x_0) := \sup\left\{ \tfrac{R+\|q\|}{\|p\|} \;\middle|\; x \in U(x_0)\right\} < \frac{R+\|q_0\|+\frac{1}{2}\|p_0\|}{\frac{1}{2}\|p_0\|}\,.$$

Therefore, for all $x \in U(x_0)$, one concludes $\Omega^+(x) =$

$$\lim_{t\to+\infty} \Phi_{-T-t} \circ \Phi_{t+T}^{(0)}(x) = \lim_{t\to+\infty} \Phi_{-T} \circ \left(\Phi_{-t} \circ \Phi_t^{(0)}\right) \circ \Phi_T^{(0)}(x) = \Phi_{-T} \circ \Phi_T^{(0)}(x),$$

which in turn implies the existence and smoothness of the Møller transformation by Theorem 3.45. The fact that the mapping Ω^+ is symplectic and thus volume preserving is a consequence of the corresponding property for Φ_t and $\Phi_t^{(0)}$, i.e., of Theorem 10.13.

• From energy conservation ($H \circ \Phi_{-t} = H$ and $H^{(0)} \circ \Phi_t^{(0)} = H^{(0)}$), one gets for $x = (p, q) \in P_+^{(0)}$ that $H \circ \Omega^\pm(x) =$

$$\lim_{t\to\pm\infty} H \circ \Phi_t^{(0)}(x) = \lim_{t\to\pm\infty} H^{(0)} \circ \Phi_t^{(0)}(x) + \lim_{t\to\pm\infty} V(q + tp) = H^{(0)}(x)\,.$$

• From Theorem 11.5 and the relation $\mathcal{T} \circ \mathcal{T} = \mathrm{Id}_P$, one has

$$\begin{aligned}\Omega^- &= \lim_{t\to+\infty} \Phi_t \circ \Phi^{(0)}_{-t} = \lim_{t\to+\infty} (\mathcal{T} \circ \Phi_{-t} \circ \mathcal{T}) \circ (\mathcal{T} \circ \Phi^{(0)}_t \circ \mathcal{T}) \\ &= \mathcal{T} \circ \left(\lim_{t\to+\infty} \Phi_{-t} \circ \Phi^{(0)}_t\right) \circ \mathcal{T} = \mathcal{T} \circ \Omega^+ \circ \mathcal{T} .\end{aligned}$$

This relation is valid whenever the limit Ω^+ exists, which we are about to prove happens in particular for all short range potentials. Therefore, $\Omega^\pm \circ \Phi^{(0)}_t = \left(\lim_{s\to\pm\infty} \Phi_{-s} \circ \Phi^{(0)}_s\right) \circ \Phi^{(0)}_t = \lim_{s\to\pm\infty} \Phi_{-s} \circ \Phi^{(0)}_{s+t} = \lim_{u\to\pm\infty} \Phi_{t-u} \circ \Phi^{(0)}_u = \Phi_t \circ \left(\lim_{u\to\pm\infty} \Phi_{-u} \circ \Phi^{(0)}_u\right) = \Phi_t \circ \Omega^\pm$.

• Similar to the case of solving the initial value problem locally in time (Theorem 3.17 by Picard-Lindelöf), we will now represent the Møller transformations as fixed points of contraction mappings. The short range condition from Definition 12.1 corresponds to a Lipschitz condition at infinity.

We assume that for $x_0 = (p_0, q_0) \in P^{(0)}_+$, there exists an initial condition $x'_0 \in P$ with $\lim_{t\to+\infty}[q(t, x'_0) - q^{(0)}(t, x_0)] = 0$ (where $(p^{(0)}, q^{(0)}) = \Phi^{(0)}$ denotes free motion, namely $q^{(0)}(t) \equiv q^{(0)}(t, x_0) = q_0 + p_0 t$).

Then it would also follow that $\lim_{t\to\infty} p(t, x'_0) = p_0$. The difference vector $Q(t) \equiv Q_{x_0}(t) := q(t, x'_0) - q^{(0)}(t, x_0)$ would satisfy the differential equation

$$Q''(t) = -\nabla V\big(q(t, x'_0)\big) ,$$

and we can rewrite it in the form

$$Q''(t) = -\nabla V\big(q^{(0)}(t) + Q(t)\big) . \tag{12.2.4}$$

Using the boundary condition $\lim_{t\to\infty} Q(t) = 0$, we are looking for solutions to the integral equation

$$Q(t) = -\int_t^\infty \int_s^\infty \nabla V\big(q^{(0)}(\tau) + Q(\tau)\big)\, \mathrm{d}\tau\, \mathrm{d}s . \tag{12.2.5}$$

Thus for all times $T \in \mathbb{R}$, the function $Q\restriction_{[T,\infty)}$ is a fixed point of the mapping $A_{x_0} \equiv A^{(T)}_{x_0}$,

$$(A_{x_0} w)(t) := -\int_t^\infty \int_s^\infty \nabla V\big(q^{(0)}(\tau) + w(\tau)\big)\, \mathrm{d}\tau\, \mathrm{d}s \qquad \big(t \in [T, \infty)\big) \tag{12.2.6}$$

on the normed space

$$C \equiv C^{(T)} := \left\{ w \in C\big([T, \infty), \mathbb{R}^d\big) \;\middle|\; \|w\|^{(T)} := \sup_{t\in[T,\infty)} \|w(t)\| < \infty \right\} .$$

By Theorem D.1 from Appendix D, $\big(C^{(T)}, \|\cdot\|^{(T)}\big)$ is a complete metric space.
• Actually, A_{x_0} maps the space C into itself, because for $c_1 := \|q_0\| + \|w\|^{(T)}$ and $T \geq c_1/\|p_0\|$, one has

$$\begin{aligned}\|A_{x_0}^{(T)}(w)\|^{(T)} &\leq \int_T^\infty \int_s^\infty \big\|\nabla V\big(q^{(0)}(\tau)+w(\tau)\big)\big\|\,\mathrm{d}\tau\,\mathrm{d}s\\ &\leq d\,c\int_T^\infty \int_s^\infty \big\langle q^{(0)}(\tau)+w(\tau)\big\rangle^{-2-\varepsilon}\,\mathrm{d}\tau\,\mathrm{d}s\\ &\leq d\,c\int_T^\infty \int_s^\infty \langle \|p_0\|\tau - c_1\rangle^{-2-\varepsilon}\,\mathrm{d}\tau\,\mathrm{d}s < \infty\,.\end{aligned}$$

The constant $c > 0$ has been taken from Definition 12.1 of short range potentials. In the last inequality, we have used that for $x_0 = (p_0, q_0) \in P_+^{(0)}$, the initial momentum p_0 is not 0.
• For large initial times T, the mapping $A_{x_0}^{(T)} : C^{(T)} \to C^{(T)}$ is also a contraction, because for $w_0, w_1 \in C^{(T)}$ one has

$$\begin{aligned}&\big\|A_{x_0}^{(T)}(w_1) - A_{x_0}^{(T)}(w_0)\big\|^{(T)}\\ &\leq \int_T^\infty \int_s^\infty \big\|\nabla V\big(q^{(0)}(\tau)+w_1(\tau)\big) - \nabla V\big(q^{(0)}(\tau)+w_0(\tau)\big)\big\|\,\mathrm{d}\tau\,\mathrm{d}s\\ &= \int_T^\infty \int_s^\infty \left\|\int_0^1 \mathrm{D}\nabla V\big(q^{(0)}(\tau)+w_r(\tau)\big)\cdot\big(w_1(\tau)-w_0(\tau)\big)\,\mathrm{d}r\right\|\mathrm{d}\tau\,\mathrm{d}s\\ &\leq c_3\int_T^\infty \int_s^\infty \langle \|p_0\|\tau - c_2\rangle^{-3-\varepsilon}\,\mathrm{d}\tau\,\mathrm{d}s\ \ \|w_1-w_0\|^{(T)}\,. \end{aligned} \tag{12.2.7}$$

In this estimate, we have used the fundamental theorem of calculus, which implies this identity, which is also called the *Hadamard lemma*: For $F(x) := \nabla V\,\big(q^{(0)}(\tau)+x\big)$, $x_i := w_i(\tau)\quad (i = 0, 1)$ and $x_r := (1-r)x_0 + r x_1$, one has:

$$F(x_1) - F(x_0) = \int_0^1 \frac{\mathrm{d}F}{\mathrm{d}r}(x_r)\,\mathrm{d}r = \int_0^1 \mathrm{D}F(x_r)\frac{\mathrm{d}x_r}{\mathrm{d}r}\,\mathrm{d}r = \int_0^1 \mathrm{D}F(x_r)\cdot(x_1-x_0)\,\mathrm{d}r.$$

In (12.2.7), one can choose $c_2 := \|q_0\| + \|w_0\|^{(T)} + \|w_1\|^{(T)}$. For $T \to +\infty$, the Lipschitz constant $c_3\int_T^\infty\int_s^\infty \langle\|p_0\|\tau - c_2\rangle^{-3\varepsilon}\,\mathrm{d}\tau\,\mathrm{d}s$ of $A_{x_0}^{(T)}$ goes to 0, so the mapping is a contraction for large T.

Thus all hypotheses of Banach's fixed point theorem (Theorem D.3) are verified, and we obtain a unique fixed point Q, initially as a curve on the interval $[T, \infty)$, then by solving the differential equation (12.2.4) as a curve $Q : \mathbb{R} \to \mathbb{R}^d$.

• Q_{x_0} depends continuously on the initial conditions $x_0 \in P_+^{(0)}$. Indeed, for given $w \in C$, the mapping $x_0 \mapsto (A_{x_0}w)(t)$ defined in (12.2.6) is continuously differentiable for all $t \in [T, \infty)$. The directional derivative $\frac{\partial}{\partial v}$ in direction $v = \binom{v_p}{v_q} \in \mathbb{R}_p^d \times \mathbb{R}_q^d$ in phase space is given in terms of $\tilde{v}(t) := v_q + t v_p$ as

$$\partial_v (A_{x_0} w)(t) = -\int_t^\infty \int_s^\infty \mathrm{D}\nabla V\big(q^{(0)}(\tau) + w(\tau)\big) \cdot \tilde{v}(\tau)\,\mathrm{d}\tau\,\mathrm{d}s\,,$$

where the exchange of differentiation and integration is justified by the estimates (locally uniform in x_0)

$$\mathrm{D}\nabla V\big(q^{(0)}(\tau) + w(\tau)\big) = \mathcal{O}\left(\langle\tau\rangle^{-3-\varepsilon}\right) \quad , \quad \tilde{v}(\tau) = \mathcal{O}\big(\langle\tau\rangle\big)$$

and Lebesgue's theorem.
This implies that $\partial_v (A_{x_0} w)(t) = \mathcal{O}(\langle t\rangle^{-\varepsilon})$, hence $\|\partial_v A_{x_0} w\|^{(T)} < \infty$. The continuity of $x_0 \mapsto Q_{x_0}$ is concluded by applying the parametrized fixed point theorem (Theorem D.4).
• The Møller transformation is therefore given by

$$\Omega^+(x_0) = \big(p_0 + Q'_{x_0}(0)\,,\, q_0 + Q_{x_0}(0)\big) \qquad \Big(x_0 = (p_0, q_0) \in P_+^{(0)}\Big)$$

and is continuous there. The existence and continuity of the inverse mapping

$$(\Omega^\pm)^{-1} : s^\pm \to P_+^{(0)} \quad , \quad x \mapsto \lim_{t\to\pm\infty} \Phi_{-t}^{(0)} \circ \Phi_t(x)$$

follows in analogous manner, by exchanging the roles of the two dynamics. □

12.13 Exercise (Møller transformations in 1D) Show in $d = 1$ dimension for a short range potential V that the *scattering transformation* $\mathcal{S} := (\Omega^+)^{-1} \circ \Omega^-$ (see (12.2.8)) for energies larger than $V_{\max} := \sup_{q\in\mathbb{R}} V(q) \geq 0$ is given by

$$\mathcal{S}(p, q) = \big(p, q - p\,\tau(p)\big) \qquad \big((p, q) \in \mathbb{R}^2\,,\; E := \tfrac{1}{2}p^2 > V_{\max}\big)\,,$$

with a *time delay*

$$\tau(p) := \int_{\mathbb{R}} \Big[\big(2(E - V(q))\big)^{-1/2} - (2E)^{-1/2}\Big]\,\mathrm{d}q\,;$$

also see THIRRING [Th1], Chapter 3.4, and our Section 12.4). ◊

12.14 Example (Decomposition of Phase Space)
Take the phase space $P = \mathbb{R}_p \times \mathbb{R}_q$ and the potential V as in Figure 12.2.2 (top part). The phase space points $x_1, x_3 \in s$ are scattering states, whereas $x_2 \in t^+$ and $x_4 \in t^-$ are trapped states. Bound states are marked in black. ◊

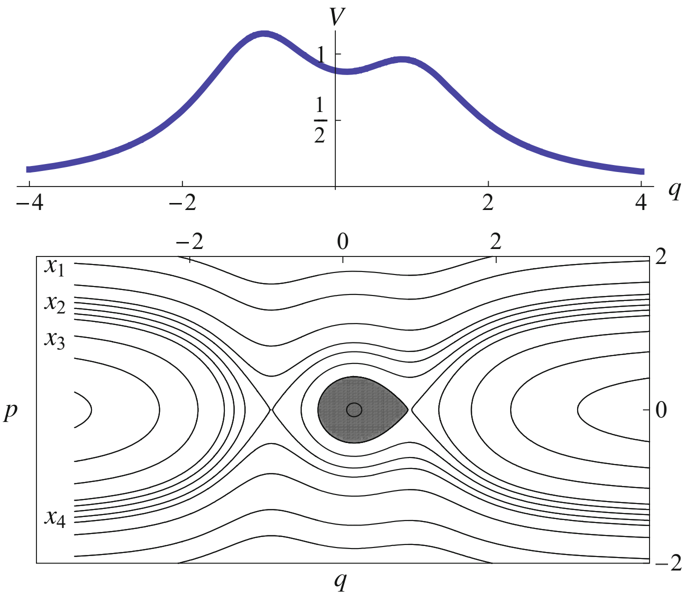


Figure 12.2.2 Bound, scattering, and trapped states (the former in black)

Intuitively speaking, the image $s^{\pm}$ of $P_+^{(0)}$ under $\Omega^{\pm}$ consists of those phase space points that escape to infinity in the future or the past respectively. In general, $s^+ \neq s^-$. A particle that is bound in the past may well be free in the future and vice versa. For instance, a meteorite could be captured by the earth-moon system. But is such an event to be expected as likely?

To answer this question, we consider the measure[6] $\lambda^{2d}(s^+ \triangle s^-)$ of the set of those states that only scatter in one time direction. Here λ^{2d} denotes the Lebesgue measure on $P = \mathbb{R}^d_p \times \mathbb{R}^d_q$.

12.15 Theorem (Measure of the Trapped States)
For scattering in potentials $V \in C^2_c(\mathbb{R}^d_q, \mathbb{R})$, the Lebesgue measure of the trapped states $t = s^+ \triangle s^-$ is zero:

$$\lambda^{2d}(t) = 0\,.$$

The intuitive reason why this theorem holds is that phase space volume arriving from infinity cannot pile up in finite space because the flow Φ_t is measure preserving. This is the contents of Schwarzschild's capture theorem:

12.16 Theorem (Schwarzschild's Capture Theorem)
Let $\Phi : \mathbb{Z} \times P \to P$ be a dynamical system that preserves a measure μ on P, and let $A \subseteq P$ be measurable with $\mu(A) < \infty$.
Then the subsets $A^{\pm} := \cap_{t \in \mathbb{N}_0} \Phi_{\pm t}(A)$ of A satisfy:

[6] The *symmetric difference* of two subsets $A, B \subset M$ is $A \triangle B := (A \setminus B) \cup (B \setminus A)$.

$$\mu(A^+) = \mu(A^+ \cap A^-) = \mu(A^-) \quad , \textit{hence} \quad \mu(A^+ \triangle A^-) = 0\,.$$

Proof: For all $T \in \mathbb{Z}$, the Φ-invariance of μ implies that

$$\mu(A^\pm) = \mu\big(\cap_{t\in\mathbb{N}_0}\Phi_{\pm t}(A)\big) = \mu\big(\Phi_T(\cap_{t\in\mathbb{N}_0}\Phi_{\pm t}(A))\big) = \mu\big(\cap_{t\in\mathbb{N}_0}\Phi_{T\pm t}(A)\big)\,,$$

hence (due to the exterior continuity of the measure μ, see BAUER [Bau], Theorem 3.2) we get $\mu(A^\pm) = \mu\big(\cap_{t\in\mathbb{Z}}\Phi_t(A)\big) = \mu(A^+ \cap A^-)$. □

Proof of Theorem 12.15. The set $t = s^+ \triangle s^-$ of trapped states is measurable because s^+ and s^- are open. t is contained in the subset $H^{-1}([0, E_0])$ of phase space, where E_0 is the energy threshold from Corollary 12.8.
For $k \in \mathbb{N}$, we define the sets $A_k := \{(p, q) \in H^{-1}([0, E_0]) \mid \|q\| \le k\}$. Being compact subsets of phase space, they have finite Lebesgue measure. We can apply to them the Schwarzschild capture theorem and obtain $\lambda^{2d}\left(A_k^+ \triangle A_k^-\right) = 0$, hence also $\lambda^{2d}\left(\cup_{k\in\mathbb{N}} A_k^+ \triangle A_k^-\right) = 0$.
On the other hand, the set t of trapped states is contained in $\cup_{k\in\mathbb{N}} A_k^+ \triangle A_k^-$; the claim follows. □

Assuming the existence of the Møller transformations and asymptotic completeness, we can define the *scattering transform*

$$\mathcal{S} : D \to D \qquad x \mapsto (\Omega^+)^{-1} \circ \Omega^-(x) \quad \text{with} \quad D := (\Omega^-)^{-1}(s)\,, \tag{12.2.8}$$

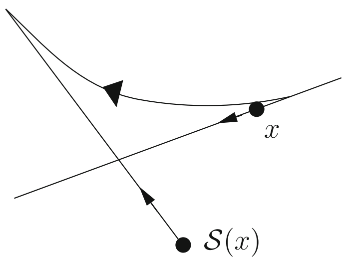
which was introduced by NARNHOFER and THIRRING in [NT], see the figure on the right. Then for $x \in D$,

$$\mathcal{S}(x) = \lim_{t\to+\infty} \Phi^{(0)}_{-t} \circ \Phi_{2t} \circ \Phi^{(0)}_{-t}(x)\,.$$

It follows from (12.2.3) that the scattering transform commutes with free motion, i.e.,

$$\mathcal{S} \circ \Phi^{(0)}_t = \Phi^{(0)}_t \circ \mathcal{S} \qquad (t \in \mathbb{R}). \tag{12.2.9}$$

All relevant information about the *result* of the scattering process are encoded in the scattering transform.



Definition of the scattering transformation $\mathcal{S}$

In contrast to the Møller transformations, the scattering transform can in principle be measured in scattering experiments, which means in mathematical modelling that it consists of time limits of observable quantities.

12.17 Remark (Quantum Mechanical Scattering)
In order to relate classical scattering results to microscopic scattering experiments, one typically has to calculate a scattering cross section from $\mathcal{S}$ (see Section 12.3).

For in microscopic scattering experiments, the impact parameter can usually not be measured.

Ironically, it is the null set t of classically trapped states that plays a central role especially in the quantum mechanical scattering process. Its so-called *resonances* can often be related to this null set. The reason for this relation is as follows: Scattering states s with a small distance to t spend a long time near bound states and have therefore a large time delay. As there is an interference effect in quantum mechanics, by which such differences in travel time lead to phase shifts with constructive or destructive interference of amplitudes in the outgoing wave, resonances can occur.

The structure of t can be rather intricate. For instance, in the n-center problem, which is relevant for scattering by molecules, this set has locally the structure of a Cantor set provided $n \geq 3$ [Kn3]. A similar phenomenon occurs in scattering by circular discs (cf. Example 2.18.3), see UZY SMILANSKY [Smi], and [KS]. ◊

12.3 The Differential Cross Section

In scattering experiments with electrons, atoms and other particles, one frequently shoots a ray of particles at a target and then measures the intensity of scattered particles in dependence on the direction. In doing so, the width of the ray is large compared to the distance of neighboring target particles, so the impact parameter cannot be measured directly.

12.18 Example (Rutherford's Experiment) Thomson, who in 1897 had discovered the *electron* by means of cathode ray experiments, designed a model of atoms called *plum pudding model*, in which the negatively charged electrons are contained in a homogeneous background of positive charge, like raisins in a pudding.

In 1909, under the guidance of Rutherford, Geiger and Marsden tested this model by aiming alpha particles at a very thin foil of gold. From the angle dependence of the scattered alpha particles, which turned out to correspond to the scattering cross section for the Coulomb potential (see Theorem 12.21), Rutherford deduced his atomic model in 1911, postulating the existence of a *nucleus*. ◊

The differential cross section is a quantity that can be measured in an experiment; it tells how the angle distribution of the intensity of scattered particles depends on energy and incoming angle. This is done under the assumption that the impact parameter of the incoming particles is equidistributed.

How can this verbal definition be formalized? First, for energy $E > 0$, let

$$A_E^{\pm} := \left\{ \left(\hat{p}^{\pm}(x), q_{\perp}^{\pm}(x) \right) \in TS^{d-1} \mid x \in s_E \right\} \tag{12.3.1}$$

be the set of scattering data for this energy.[7] As there is exactly one scattering orbit for every point in $A_E^{\pm}$, we can define a continuous mapping by

$$\varphi_E : A_E^- \to S^{d-1} \quad , \quad \left(\hat{p}^-(x), q_\perp^-(x)\right) \mapsto \hat{p}^+(x)\,.$$

Here x is an arbitrary point on the scattering orbit. By Corollary 12.8 for long range potentials and energies $E > 0$, we have $A_E^{\pm} = TS^{d-1}$.

12.19 Definition

- For $\theta^- \in S^{d-1}$, let λ_{θ^-} denote the $(d-1)$-dimensional Lebesgue measure on the vector space $T_{\theta^-}S^{d-1} := \{q \in \mathbb{R}^d \mid \langle q, \theta^- \rangle = 0\}$. The image measure of λ_{θ^-} under the mapping
$$\varphi_{E,\theta^-} := \varphi_E\restriction_{T_{\theta^-}S^{d-1}} : T_{\theta^-}S^{d-1} \to S^{d-1}$$
will be denoted as $\sigma(E, \theta^-)$.
- The **differential cross section** $\frac{d\sigma}{d\theta}(E, \theta^-, \theta^+)$ is the density of this image measure at the point $\theta^+ \in S^{d-1}$.

12.20 Remark Speaking more precisely, the differential cross section is the Radon-Nikodym derivative of the measure $\sigma(E, \theta^-)$ on the sphere, with respect to the rotationally invariant probability measure μ on S^{d-1}. As the image of the Lebesgue measure, $\sigma(E, \theta^-)$ is not a finite measure. Whether it is absolutely continuous with respect to μ, as required in the theorem of Radon-Nikodym, needs to be checked for each particular scattering problem. $\diamondsuit$

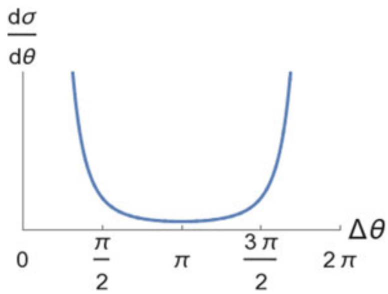
12.21 Theorem (Rutherford Scattering Cross Section)
We consider scattering by the Coulomb potential

$$V \in C^\infty\left(\widehat{M}, \mathbb{R}\right) \quad , \quad V(q) = -\frac{Z}{\|q\|}$$

on the configuration space $\widehat{M} := \mathbb{R}^d \setminus \{0\}$ *for* $d \geq 2$ *with a charge* $Z \in \mathbb{R} \setminus \{0\}$. *Then for* $E > 0$, *the differential cross section equals*

$$\frac{d\sigma}{d\theta^+}(E, \theta^-, \theta^+) = \left(\frac{|Z|}{4E\sin^2\left(\frac{1}{2}\Delta\theta\right)}\right)^{d-1},$$

where $\Delta\theta \in (0, \pi]$ *denotes the angle between* θ^- *and* θ^+.



[7] If E is a regular value of H, one can show (analogous to Theorem 12.15) that the complement sets $TS^{d-1} \setminus A_E^{\pm}$ have measure zero, in reference to the natural measure on TS^{d-1}. These sets belong to the trapped orbits.

Proof:
• We first consider the case of $d = 2$ dimensions.

For a value ℓ of the angular momentum L, the impact parameter at energy E equals $q_\perp^- = \frac{\ell}{\|p^-\|} = \frac{\ell}{\sqrt{2E}}$. We now calculate the dependence of the scattering angle

$$\Delta\theta(E,\ell) = 2\int_{r_{\min}}^{\infty} \frac{\mathrm{d}\varphi}{\mathrm{d}r}\,\mathrm{d}r + \pi\,, \tag{12.3.2}$$

on the angular momentum in polar coordinates (r, φ); here $r_{\min}$ denotes the pericenter radius, which is the radius of the point on the hyperbola that is closest to the origin. At this point, the radial velocity is 0, hence

$$E - W_\ell(r) = E + \frac{Z}{r} - \frac{\ell^2}{2r^2} = 0\,,$$

and this determines $r_{\min} \geq 0$. Using

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{\ell}{r^2} \quad \text{and} \quad \frac{\mathrm{d}r}{\mathrm{d}t} = \sqrt{2(E - W_\ell(r))}\,,$$

see the formula (1.6) in the introduction, one obtains (modulo 2π):

$$\Delta\theta(E,\ell) = 2\int_{r_{\min}}^{\infty} \frac{\ell^2}{r^2\sqrt{2\left(E + \frac{Z}{r} - \frac{\ell^2}{2r^2}\right)}}\,\mathrm{d}r = 2\tan^{-1}\left(\frac{Z}{\ell\sqrt{2E}}\right),$$

or

$$q_\perp^- = \frac{\ell}{\sqrt{2E}} = \frac{Z}{2E}\cot\left(\frac{\Delta\theta}{2}\right). \tag{12.3.3}$$

The derivative with respect to $\Delta\theta$ is, in absolute value,

$$\left|\frac{\mathrm{d}q_\perp^-}{\mathrm{d}\Delta\theta}\right| = \frac{|Z|}{4E}\frac{1}{\sin^2\left(\frac{1}{2}\Delta\theta\right)}\,,$$

which is the claim for the case $d = 2$.
• For $d \geq 3$, notice that in the vector space $T_{\theta^-}S^{d-1} \cong \mathbb{R}^{d-1}$ of impact parameters (see Definition 12.19), the ball of radius r has volume $v_{d-1}\int_0^r \tilde{r}^{d-2}\,\mathrm{d}\tilde{r}$, whereas the segment of the sphere S^{d-1} that is defined by the condition $\Delta\tilde{\theta} \in [0, \Delta\theta]$ has volume

$$v_{d-1}\int_0^{\Delta\theta} (\sin(\Delta\tilde{\theta}))^{d-2}\,\mathrm{d}\Delta\tilde{\theta}\,.$$

Here v_d denotes the volume of the d-dimensional unit ball. The ratio of both integrands is the differential cross section. For $\tilde{r}(\Delta\tilde{\theta}) := \|q_\perp^-(\Delta\tilde{\theta})\|$, the two variables of integration are, by (12.3.3), related as $\tilde{r}(\Delta\tilde{\theta}) = \frac{|Z|}{2E}\cot\left(\frac{\Delta\tilde{\theta}}{2}\right)$. Therefore, using the trigonometric formula $\sin(\Delta\tilde{\theta}) = 2\sin\left(\frac{\Delta\tilde{\theta}}{2}\right)\cos\left(\frac{\Delta\tilde{\theta}}{2}\right)$, one concludes that

$$\frac{v_{d-1}\tilde{r}^{d-2}\frac{\mathrm{d}\tilde{r}}{\mathrm{d}\tilde{\theta}}}{v_{d-1}\left(\sin(\Delta\tilde{\theta})\right)^{d-2}} = \left(\frac{|Z|\cot\left(\frac{\Delta\tilde{\theta}}{2}\right)}{2E\sin\Delta\tilde{\theta}}\right)^{d-2}\frac{|Z|}{4E\sin^2\left(\frac{\Delta\tilde{\theta}}{2}\right)} = \left(\frac{|Z|}{4E\sin^2\left(\frac{\Delta\tilde{\theta}}{2}\right)}\right)^{d-1},$$

which proves the claimed formula for dimension $d \geq 3$. □

12.22 Remarks (Rutherford Scattering)

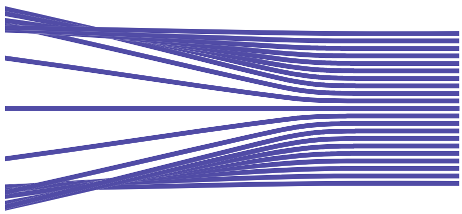
1. Whereas the differential cross section diverges in forward direction (because far distant particles experience little deflection), it is also nonzero for the backwards direction, which corresponds to a collision orbit.[8] This was incompatible with the plum pudding model.
2. Note that the Rutherford cross section is independent of the sign of the charge Z.
3. The analogous cross section in quantum mechanics is of the same form. This was fortunate because in 1911, when Rutherford derived his cross section from classical mechanics, the Schrödinger equation of quantum mechanics was not available yet. ◊

12.23 Example (Rainbow Singularity) Let $V \in C^2(\mathbb{R}^2, \mathbb{R})$ be a centrally symmetric potential with compact support, i.e., $V(q) = W(\|q\|)$ for some function W satisfying $W(r) = 0$ for all sufficiently large r, say $r \geq R > 0$.

As V is bounded, there are energies $E > \sup_q V(q)$, and it is for those that we want to analyze the cross section. For impact parameters whose absolute value is larger than R, the orbit does not hit the support of V and is therefore not deflected at all. Likewise, every orbit with impact parameter 0 is a straight line, for reasons of symmetry.

On the other hand, for $V \neq 0$, there do exist orbits that are deflected (think why). Since the dependence of the deflection $\Delta\theta$ on the impact parameter is continuously differentiable and we have the symmetry $\Delta\theta(-q_\perp) = -\Delta\theta(q_\perp)$, there must exist a nonzero maximum of $\Delta\theta$. At this maximum, one has $\frac{\mathrm{d}\Delta\theta}{\mathrm{d}q_\perp} = 0$, so for this angle, the differential cross section diverges, in contrast to the cross section for the singular Coulomb potential!

For instance, this phenomenon occurs for the scattering of light in water droplets and is called *rainbow singularity* for this reason. This singularity is a convolution singularity in the sense of Example 8.39. ◊



[8] Strictly speaking, the Kepler motion had to be regularized for these collision orbits. Then the expression for the Rutherford cross section is obtained also for $\Delta\theta = \pi$.

12.24 Exercise (Scattering at High Energies) Let the support of the potential $V \in C^2(\mathbb{R}^d, \mathbb{R})$, $V \neq 0$ be contained in a ball of radius R, and let

$$\|V\|_\infty := \sup_{q\in\mathbb{R}^d} |V(q)| \quad , \quad \|\nabla V\|_\infty := \sup_{q\in\mathbb{R}^d} \|\nabla V(q)\| \,.$$

(a) Show that for all $E > \|V\|_\infty$, with the maximum and minimum speeds

$$v_{\max} := \sqrt{2(E + \|V\|_\infty)} \quad , \quad v_{\min} := \sqrt{2(E - \|V\|_\infty)} \,,$$

the rate of change of direction for a particle with energy E in the potential can be estimated by

$$\left\|\frac{\mathrm{d}\theta}{\mathrm{d}t}\right\| \le \frac{\|\nabla V\|_\infty}{v_{\min}} \,.$$

(b) Show furthermore that for times $t \in \left[0, \frac{2v_{\min}}{\|\nabla V\|_\infty}\right]$, the distance traveled by the particle satisfies the two-sided estimate

$$0 \le t\left(v_{\min} - \tfrac{1}{2}\|\nabla V\|_\infty\, t\right) \le \|q(t) - q(0)\| \le v_{\max}\, t \,.$$

(c) Show that for particle energies $E \ge \|V\|_\infty + 2R\,\|\nabla V\|_\infty$, the particle is scattered and undergoes a deflection of at most

$$\|\Delta\theta\| \le \frac{2R\,\|\nabla V\|_\infty}{E - \|V\|_\infty} \le 1 \,.$$

(Generally, at large energies E, smooth long range potentials lead to scattering with small deflections $\mathcal{O}(1/E)$.) ◊

12.4 Time Delay, Radon Transform, and Inverse Scattering Theory

How do we actually know how the interactions between microscopic particles look? The example of the Rutherford atomic model from Section 12.3 indicates that this knowledge can be obtained from scattering experiments. However, the following questions arise:

1. Could a different potential than Coulomb also lead to Rutherford's cross section (Theorem 12.21)?
2. If Rutherford had not started with the correct formula for the potential, could he instead have calculated it, in principle, from his scattering data?
3. Which quantities that are accessible to measurement, if any, allow the reconstruction of the potential?

Of course, if the first question had a positive answer, this would make it impossible for an inverse scattering method, as addressed in the following questions, to exist.

12.25 Examples

1. **Nonexistence of a 1D Inverse Scattering Theory**
   This is indeed what happens in $d = 1$ dimension: Let $V : \mathbb{R} \to \mathbb{R}$ be a short range potential and Φ the corresponding flow.

   Then if we consider the potential V shifted by $\ell \in \mathbb{R}$, i.e., the potential $V^{(\ell)}(q) := V(q - \ell)$, and transform the phase space correspondingly by

$$W^{(\ell)} : P \to P \quad , \quad (p, q) \mapsto (p, q + \ell)\,,$$

   then the flow $\Psi^{(\ell)}$ corresponding to $V^{(\ell)}$ is of the form

$$\Psi_t^{(\ell)} = W^{(\ell)} \circ \Phi_t \circ W^{(-\ell)} \qquad (t \in \mathbb{R}).$$

   We now consider the phase space domain $P^+ := H^{-1}\big((V_{\max}, \infty)\big)$ with $V_{\max} := \sup_{q \in \mathbb{R}} V(q)$. In this domain, solutions $\big(p(t, x_0), q(t, x_0)\big) := \Phi_t(x_0)$ do not reverse direction, i.e., $\operatorname{sign}\big(p(t, x_0)\big)$ is independent of t.
   Let $\mathcal{S}^{(\ell)}$ be the scattering transform for $\Psi^{(\ell)}$ and $\mathcal{S}$ the one for Φ. Then

$$\mathcal{S}^{(\ell)}(x) = \mathcal{S}(x) \qquad (x \in P^+), \tag{12.4.1}$$

   and one cannot see from the scattering data that the potential was shifted.

   Equation (12.4.1) follows for $x = (p, q)$ from the relation

$$W^{(\ell)}(x) = (p, q + \ell) = \Phi^{(0)}_{\ell/p}(x) \ :$$

$$\begin{aligned} \mathcal{S}^{(\ell)}(x) &= \lim_{t\to\infty} \Phi^{(0)}_{-t} \circ \Psi^{(\ell)}_{2t} \circ \Phi^{(0)}_{-t}(x) \\ &= \lim_{t\to\infty} \Phi^{(0)}_{-t} \circ W^{(\ell)} \circ \Phi_{2t} \circ W^{(-\ell)} \circ \Phi^{(0)}_{-t}(x) \\ &= W^{(\ell)} \circ \Big(\lim_{t\to\infty} \Phi^{(0)}_{-t} \circ \Phi_{2t} \circ \Phi^{(0)}_{-t}\Big) \circ W^{(-\ell)}(x) \\ &= W^{(\ell)} \circ \mathcal{S} \circ W^{(-\ell)}(x) = \Phi^{(0)}_{\ell/p} \circ \mathcal{S} \circ \Phi^{(0)}_{-\ell/p}(x) = \mathcal{S}(x)\,. \end{aligned}$$

   In the last step, we used (12.2.9), namely that $\mathcal{S}$ commutes with $\Phi^{(0)}$. (12.4.1) can also be obtained by explicit calculation (Exercise 12.13).
2. **Differential Cross Section** It is impossible in any space dimension d to reconstruct the potential V from the differential cross section (Definition 12.19), because the latter is invariant under translation of V. ◊

Fortunately, the situation is better in dimensions $d \geq 2$ and for other kinds of scattering data. Different potentials lead to different scattering transforms, and there also exist methods to reconstruct the potentials from scattering data like the time delay.

These methods are related to those of the *Radon* transform or the *X-ray transform*, which is used in computer tomography.

12.26 Definition In dimension $d \geq 2$, the **X-ray transform** of a function $f \in C^2(\mathbb{R}^d, \mathbb{R})$ with compact support is the function (see Definition 12.1.16)

$$\mathcal{R}f : TS^{d-1} \to \mathbb{R} \quad , \quad (u, v) \mapsto \int_{\mathbb{R}} f(v + tu)\, \mathrm{d}t\, .$$

12.27 Remarks (X-Ray and Radon Transforms)

1. In the X-ray transform, f is integrated over that straight line with direction $u \in S^{d-1}$ which comes closest to the origin $0 \in \mathbb{R}^d$ in the point $v \in T_u S^{d-1} = \{q \in \mathbb{R}^d \mid \langle q, u \rangle = 0\}$.
2. In the Radon transform[9] one integrates over hypersurfaces rather than over straight lines. So for $d = 2$, the Radon transform essentially coincides with the X-ray transform.
3. Computer tomography was developed theoretically and practically by the physicist Allan Cormack (1924–1998) and the engineer Godfrey Hounsfield (1919–2004). It was in the 1970s, at a time when effective calculation of the inverse Radon transform by computers became feasible, that the first prototypes came into existence.
4. A standard textbook on computer tomography is [Nat] by NATTERER.

   The book [Hel] by HELGASON on Radon transforms (as defined more generally in terms of group theory and thus including the X-ray transform) is also available on its author's home page.
5. It is subtle to determine precisely the classes of functions f to which the Radon or X-ray transform can be applied, but this question is of practical importance. Think of $f : \mathbb{R}^3 \to \mathbb{R}^+$ as the optical density of the tissue being X-rayed (in the wavelength range of X-rays), a quantity called the *X-ray absorption factor*). Then the value of f will have a jump discontinuity at the interface of bone and muscle tissue (see Chapter IV.2 in [Nat]). $\Diamond$

In computer tomography, one measures the intensities of the X-rays passing through the object, which means one measures $\mathcal{R}f$. The mathematical task consists of reconstructing f from $\mathcal{R}f$. To do this, the first important observation is:

- $\mathcal{R}$ is a linear mapping. So the question is whether its kernel consists only of the zero function.

[9] The Austrian mathematician *Johann Radon* (1887–1956) investigated this integral tranform named after him in a 1917 paper.

- For the domains of $\mathcal{R}f$ and of f, one has $\dim(TS^{d-1}) = 2d-2$, which for $d \geq 2$ is larger or equal to $\dim(\mathbb{R}^d)$. So — in contrast to one dimension — we have at least a chance to solve the problem in higher dimensions.

The key for inverting the operator $\mathcal{R}$ is the following statement, called *Fourier slice theorem*.

- Here, we denote by

$$\mathcal{F}_d : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d) \quad , \quad (\mathcal{F}_d f)(k) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-\mathrm{i}\,k\cdot x}\,\mathrm{d}x$$

  the Fourier transformation on the space of *Schwartz functions*[10] on $\mathbb{R}^d$.
- Moreover, for a direction $u \in S^{d-1}$, let

$$\mathcal{R}^u : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(T_u S^{d-1}) \quad , \quad (\mathcal{R}^u f)(v) = (\mathcal{R}f)(u, v)$$

  be the restriction of the X-ray transform (namely what the computer tomograph sees when X-rays pass in direction u).
- The Fourier transformation of a function on the $(d-1)$-dimensional subspace $T_u S^{d-1} = \{v \in \mathbb{R}^d \mid \langle u, v\rangle = 0\}$ will be denoted as $\mathcal{F}^u_{d-1}$.

12.28 Theorem (Fourier-slice-Theorem) *For $d \geq 2$, one has*

$$\left(\mathcal{F}^u_{d-1}\mathcal{R}^u f\right)(k) = \sqrt{2\pi}\,(\mathcal{F}_d f)(k) \qquad \left(u \in S^{d-1}, k \in T_u S^{d-1}\right).$$

Proof: Given $u \in S^{d-1}$, we can uniquely write any $x \in \mathbb{R}^d$ in the form $x = tu + v$ with $v \in T_u S^{d-1}$ and $t \in \mathbb{R}$. Then we obtain

$$\begin{aligned}
(2\pi)^{\frac{d-1}{2}} \left(\mathcal{F}^u_{d-1}\mathcal{R}^u f\right)(k) &= \int_{T_u S^{d-1}} (\mathcal{R}^u f)(u, v) e^{-\mathrm{i}\,v\cdot k}\,\mathrm{d}v \\
&= \int_{T_u S^{d-1}} \left(\int_{\mathbb{R}} f(v + tu)\,\mathrm{d}t \right) e^{-\mathrm{i}\,v\cdot k}\,\mathrm{d}v = \int_{T_u S^{d-1}} \int_{\mathbb{R}} f(v + tu) e^{-\mathrm{i}\,(v+tu)\cdot k}\,\mathrm{d}t\,\mathrm{d}v \\
&= \int_{\mathbb{R}^d} f(x) e^{-\mathrm{i}\,x\cdot k}\,\mathrm{d}x = (2\pi)^{d/2} (\mathcal{F}_d f)(k),
\end{aligned}$$

because $u \cdot k = 0$. □

As a consequence, the operator $\mathcal{R}$ of the X-ray transformation is invertible, because the Fourier transformation $\mathcal{F}_d : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ is invertible.

---

[10] **Definition:** The *Schwartz space* $\mathcal{S}(\mathbb{R}^d)$ is the function space

$$\mathcal{S}(\mathbb{R}^d) := \left\{ f \in C^\infty(\mathbb{R}^d, \mathbb{C}) \mid \forall\, \alpha, \beta \in \mathbb{N}_0^d : t \mapsto t^\alpha \partial^\beta f(t) \text{ is bounded} \right\}.$$

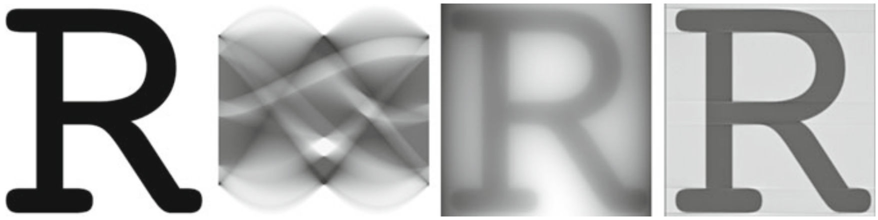


Figure 12.4.1 From left: letter R; its Radon transform (abscissa: angular variable); numerical inverse Radon transform; inverse Radon transform using a filter

12.29 Remark (Ill-Posed Problems) Integration averages out high frequency oscillations, whereas differentiation amplifies them. This results in the fact that the X-ray transformation, defined on appropriate function spaces, is a compact operator, whereas its inverse is unbounded. In practice, this leads to a strong amplification of noise in the image data, and to artefacts (see Figure 12.4.1).
So to achieve a regularization of the problem, one applies methods from the *theory of ill-posed problems*, see for instance the book [Lou] by LOUIS. ◊

Let us return to scattering theory. From the scattering transform $\mathcal{S} : D \to D$ for short range potentials defined in (12.2.8), one can calculate the transformation of the asymptotic momenta and impact parameters introduced in Theorem 12.5, because if

$$(p^+, q^+) := \mathcal{S}\big((p^-, q^-)\big) \qquad \big((p^-, q^-) \in D\big)\,,$$

we already have the asymptotic momenta $p^\pm$, and then the asymptotic impact parameters are obtained as

$$q_\perp^\pm = q^\pm - \frac{\langle q^\pm, p^\pm\rangle}{\|p^\pm\|^2}\, p^\pm\,.$$

Next to these $2d-1$ coordinates, there is another piece of information contained in $\mathcal{S}$, namely the *time delay*

$$\tau : D \to \mathbb{R} \quad , \quad \tau\big((p^-, q^-)\big) = \frac{\langle p^-, q^-\rangle - \langle p^+, q^+\rangle}{\|p^\pm\|^2}\,, \tag{12.4.2}$$

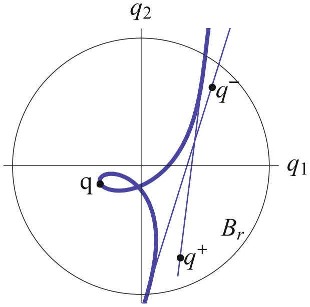
introduced by H. NARNHOFER in [Nar].[11]

12.30 Remark (Time Delay) The time delay tells how much longer (in the limit $R \to \infty$) the orbit with initial condition

$$x = (p, q) = \Omega^+\big((p^+, q^+)\big) = \Omega^-\big((p^-, q^-)\big)$$

[11] See also the survey article [CN] on classical and quantum mechanical time delays.

stays in a ball B_R^d of large radius R than orbits of free motion would do. This can be studied in the following figure.



- As the two orbits of free motion are straight lines, their intersections with the ball are segments of lengths

$$L^{\pm}(R) = 2\sqrt{R^2 - \|q_\perp^\pm\|^2} = 2R\sqrt{1 - \left(\|q_\perp^\pm\|/R\right)^2} = 2R + \mathcal{O}(1/R). \quad (12.4.3)$$

  So in the limit $R \to \infty$, the difference between L^+ and L^- tends to 0.
- For the path from $q^\pm$ to the pericenter point $q_\perp^\pm$ (where $\langle p^\pm, q_\perp^\pm \rangle = 0$), free motion takes the time

$$t^\pm := -\frac{\langle p^\pm, q^\pm \rangle}{\|p^\pm\|^2}, \quad (12.4.4)$$

  because $\langle p^\pm, q^\pm + t^\pm p^\pm \rangle = 0$. From the pericenter, the time needed yet to leave the ball B_R has absolute value $\frac{L^\pm(R)}{2\|p^\pm\|}$.
- If the true orbits and their free asymptotes already coincide outside B_R, then the time T_R which the orbit spends in B_R is exactly

$$T_R = \left(\frac{L^+(R)}{2\|p^+\|} + t^+\right) - \left(-\frac{L^-(R)}{2\|p^-\|} + t^-\right), \quad (12.4.5)$$

  otherwise the difference of T_R and this expression tends to 0 in the limit $R \to \infty$.
- Plugging (12.4.3) and (12.4.4) into (12.4.5), one obtains with τ from Formula (12.4.2):

$$\lim_{R\to\infty} \left(T_R - L^\pm(R)/\|p^\pm\|\right) = \tau\big((p^-, q^-)\big). \qquad \diamondsuit$$

12.31 Theorem (**Time Delay**) *For the orbit curve $t \mapsto q(t) \equiv q(t, x)$ with initial value $x = \Omega^\pm\big((p^\pm, q^\pm)\big) \in s$ and energy $E := H(x)$, one has*

$$\tau\big((p^-, q^-)\big) = (2E)^{-1} \int_{\mathbb{R}} \Big[2V\big(q(t)\big) + \big\langle q(t), \nabla V\big(q(t)\big)\big\rangle\Big]\, \mathrm{d}t\,. \quad (12.4.6)$$

Proof:
• By Definition (12.4.2), one has $\tau\big((p^-,q^-)\big) = \frac{\langle p^-,q^-\rangle - \langle p^+,q^+\rangle}{2E}$, because $\|p^\pm\|^2 = 2E$.
• On the other hand, by Definition (12.2.1) of the Møller transforms $\Omega^\pm$, one has

$$\lim_{t\to\pm\infty}\left[\langle p(t),q(t)\rangle - \langle p^\pm, q^\pm + tp^\pm\rangle\right] = 0$$

for $p = \frac{\mathrm{d}q}{\mathrm{d}t}$, and therefore by the fundamental theorem of calculus (as in (12.1.4)),

$$\begin{aligned}\langle p^-,q^-\rangle - \langle p^+,q^+\rangle &= \int_{\mathbb{R}}\left(2E - \frac{\mathrm{d}}{\mathrm{d}t}\langle p(t),q(t)\rangle\right)\mathrm{d}t\\ &= \int_{\mathbb{R}}\Big[2E - 2\big(E - V(q(t))\big) + \big\langle q(t), \nabla V\big(q(t)\big)\big\rangle\Big]\,\mathrm{d}t\,.\end{aligned}$$

The formula follows. □

Rather than defining the time delay as a function defined on $D \subset P^{(0)}$ as in (12.4.2), we can also fix the energy E and view the time delay as a function on TS^{d-1}.

Using the subset $A_E^- \subseteq TS^{d-1}$ from (12.3.1) associated with scattering data for energy $E > 0$, we obtain a well-defined mapping

$$\tilde\tau_E : A_E^- \to \mathbb{R} \quad,\quad \tilde\tau_E\big(\hat p^-, q_\perp^-\big) := \tau\big((p^-, q_\perp^-)\big)\,.$$

For all energies $E > E_0$, Corollary 12.8 tells us that the energy shell consists only of scattering orbits, and then $\tilde\tau_E$ is defined on all of TS^{d-1}.

We now consider a potential $V \in C^\infty(\mathbb{R}^d,\mathbb{R})$ with compact support and the function

$$f_E : \mathbb{R}^d \to \mathbb{R} \quad,\quad f_E(q) := \frac{2V(q) + \langle q, \nabla V(q)\rangle}{2E\sqrt{2(E - V(q))}}\,.$$

A Taylor estimate of $f_E(q)$ yields, uniformly in $q \in \mathbb{R}^d$, that

$$f_E(q) = \frac{2V(q) + \langle q, \nabla V(q)\rangle}{(2E)^{3/2}} + \mathcal{O}\big(E^{-5/2}\big)\,. \qquad (12.4.7)$$

So for large energies, f_E in (12.4.6) may be integrated approximately along the free orbit to obtain the time delay:

12.32 Theorem (Inverse Scattering Theory) *For energy $E > E_0$, the difference between the time delay $\tilde\tau_E$ and the X-ray transform of f_E, i.e.,*

$$\Delta_E := \tilde\tau_E - \mathcal{R}f_E :\ TS^{d-1} \to \mathbb{R}\,,$$

is of order $\sup_x |\Delta_E(x)| = \mathcal{O}\left(E^{-5/2}\right)$.

12.33 Remark (Inverse Scattering Theory)
So by measuring $\tilde{\tau}_E$, we can reconstruct the X-ray transform of f_E up to a small error of relative order $\mathcal{O}(1/E)$ for large energies E. By the Fourier Slice Theorem 12.28, f_E itself can therefore be measured in the experiment.[12]

The same also applies to a short range potential V: As can be checked by plugging in, for any dimension d, one can reconstruct V from the numerator $F(q) := 2V(q) + \langle q, \nabla V(q)\rangle$ in (12.4.7) by

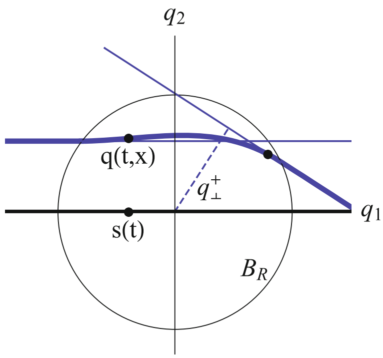
$$V(0) = \tfrac{1}{2}F(0) \quad , \quad V(q) = -\int_1^\infty F(tq)\, t\, \mathrm{d}t \quad \big(q \in \mathbb{R}^d \setminus \{0\}\big) .$$

This solution to the quasilinear partial differential equation $2V(q) + \langle q, \nabla V(q)\rangle = F(q)$ is found by integrating along the characteristic vector field. See Arnol'd [Ar3, Chap. 2.7], and Schmitz [Schm]. $\Diamond$

Proof:
- Rather than by its time t, we will parametrize the orbit curve $t \mapsto q(t,x)$ by its projection

$$s(t) := \big\langle \tilde{p}^-(x), q(t,x)\big\rangle \qquad (t \in \mathbb{R}),$$

see the figure to the right. This is possible if E is sufficiently large, because in this case, by Exercise 12.24.c, the deflection $\sphericalangle\big(\hat{p}^-(x), \frac{\mathrm{d}}{\mathrm{d}t}q(t,x)\big)$ is smaller than $\pi/2$, hence $s'(t) > 0$.



- Using

$$s'(t) = \big\langle \hat{p}^-(x), p(t,x)\big\rangle = \Big\langle \hat{p}^-(x), \tfrac{p(t,x)}{\|p(t,x)\|}\Big\rangle \sqrt{2\big(E - V(q(t,x))\big)}$$

and $\Delta\theta(t) := \sphericalangle\big(\hat{p}^-(x), p(t,x)\big) = \mathcal{O}(1/E)$ (Exercise 12.24), hence

$$\Big\langle \hat{p}^-(x), \tfrac{p(t,x)}{\|p(t,x)\|}\Big\rangle = \cos\big(\Delta\theta(t)\big) = \sqrt{1 - \sin^2(\Delta\theta(t))} = 1 + \mathcal{O}\big(E^{-2}\big) , \qquad (12.4.8)$$

we obtain

$$\tau(x) = \int_{\mathbb{R}} f_E\big(\tilde{q}(s)\big)\, \mathrm{d}s \;+ \mathcal{O}\big(E^{-5/2}\big)$$

with $\tilde{q}(s) := q\big(t(s)\big)$. Choosing $R > 0$ large enough to ensure $\mathrm{supp}(V) \subseteq B_R^d$, and thus also $\mathrm{supp}(f_E) \subseteq B_R^d$, one obtains

$$\int_{\mathbb{R}} f_E\big(\tilde{q}(s)\big)\, \mathrm{d}s = \int_{-R}^{R} f_E\big(\tilde{q}(s)\big)\, \mathrm{d}s = \int_{-R}^{R} f_E\big(q_\perp^- + s\, \hat{p}^-\big)\, \mathrm{d}s + \mathcal{O}\big(E^{-5/2}\big) .$$

[12] Up to an error depending on E, which can be found in the theorems in [Nat], Chapter IV.2.

The latter estimate follows from $\left\|\tilde{q}(s) - \left(q_\perp^- + s\hat{p}^-\right)\right\| = \mathcal{O}(1/E) \quad (|s| \le R)$ and the Taylor formula for the difference of the terms under the integral. Therefore,

$$\tilde{\tau}_E\left(\hat{p}^-, q_\perp^-\right) = \tau\left(p^-, q^-\right) = \mathcal{R}f_E\left(\hat{p}^-, q_\perp^-\right) + \mathcal{O}\left(E^{-5/2}\right), \tag{12.4.9}$$

as claimed. □

12.5 Kinematics of the Scattering of *n* Particles

We consider the motion of finitely many particles that may exercise forces upon each other, but are not subject to external forces, which would depend on the position of their center of mass.

One question is about the long term dynamics: Will the particles end up grouping together in *clusters*, with strong interaction inside the clusters, but such that the interaction between particles of different clusters decays with time?

12.34 Example (Billard Balls in Space)
If we consider n balls of radius $R > 0$ in $\mathbb{R}^d$ whose masses are $m_1, \ldots, m_n > 0$, then their configuration space is of the form

$$M := \mathbb{R}^{nd} \backslash \Delta \quad \text{with} \quad \Delta := \bigcup_{1 \le i < j \le n} \Delta_{i,j}$$

and $\Delta_{i,j} := \{q = (q_1, \ldots, q_n) \in (\mathbb{R}^d)^n \mid \|q_i - q_j\| < 2R\}$, because the midpoints of the i^{th} and the j^{th} ball must have distance at least $2R$.

According to the Hamiltonian [13]

$$H : P \to \mathbb{R} \quad , \quad H(p, q) := \sum_{i=1}^{n} \frac{\|p_i\|^2}{2m_i}$$

on the phase space $P := \mathbb{R}^{nd} \times M$ of the n particles, these particles move at constant velocity $\dot{q}_i(t) = p_i(0)/m_i$ until they hit the boundary of their configuration space.

Here we encounter clusters for the first time, because for $q \in \partial M$, we can define a graph on the set $N := \{1, \ldots, n\}$ of vertices by connecting those elements $i < j$ with an edge $\{i, j\}$ for which $\|q_i - q_j\| = 2R$ (i.e., for which the balls numbered i and j are in contact). So the set N is decomposed into clusters of vertices that are connected by edges, see Figure 12.5.1.

- We say a configuration $q \in \partial M$ has only *double collisions*, if the clusters have size at most two. Then for cluster $\{i, j\}$, the motion will be continued after the collision according to the following rules for an *elastic collision*:

[13] A more realistic study of the mechanics of billiard can be found in § 27 of SOMMERFELD [Som]. It includes the study of top and bottom spin, follow shots etc.

Let the momenta of the balls immediately before the collision be p_i^-, p_j^-, and after the collision p_i^+, p_j^+. Then we have the following conserved quantities:

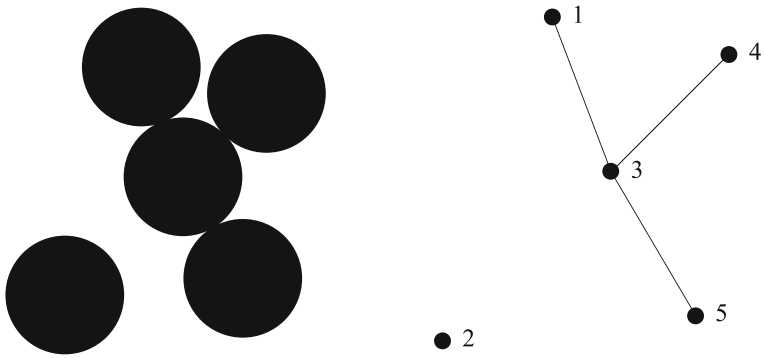


Figure 12.5.1 Configuration of billiard balls (left) with its corresponding graph (right) and its cluster decomposition {{1, 3, 4, 5}, {2}}

(a) The *total momentum* is conserved, i.e., $p_i^+ + p_j^+ = p_i^- + p_j^-$.
(b) The *total energy* is conserved, i.e., $\frac{\|p_i^+\|^2}{2m_i} + \frac{\|p_j^+\|^2}{2m_j} = \frac{\|p_i^-\|^2}{2m_i} + \frac{\|p_j^-\|^2}{2m_j}$.
(c) Decomposing the momenta into components parallel and orthogonal to the line that connects the centers of the balls, the orthogonal components are conserved, i.e., for the decomposition

$$p_k^{\pm} = p_k^{\|,\pm} + p^{\perp,\pm} \quad \text{one has} \quad p_k^{\perp,+} = p_k^{\perp,-} \qquad (k = i, j).$$

These rules imply a quadratic equation for the component of the momenta that is parallel to $q_i - q_j$. This equation has (next to the unphysical solution $p_k^{\|,+} = p_k^{\|,-}$) the solution

$$p_i^{\|,+} = \frac{m_i - m_j}{m_i + m_j}\, p_i^{\|,-} + \frac{2m_i}{m_i + m_j}\, p_j^{\|,-} \quad , \quad p_j^{\|,+} = \frac{m_j - m_i}{m_i + m_j}\, p_j^{\|,-} + \frac{2m_j}{m_i + m_j}\, p_i^{\|,-} \,. \tag{12.5.1}$$

The solution can be continued with these data.

- However, if, as in Figure 12.5.1, three or more balls collide simultaneously in a cluster, there is in general no rule for continuing the motion that would be continuous in the initial data. So one assumes such initial conditions are given that do not lead to this kind of collision. This is indeed the case except on a null set of exceptional initial conditions.

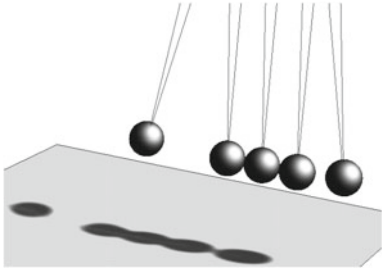
In 1998, BURAGO, FERLEGER and KONONENKO showed in [BFK] that, under these hypotheses, there will be only finitely many collisions, and they also gave an upper bound for the number of collisions.

As the momenta remain unchanged after the last collision, we obtain a decomposition into groups of balls that move with the same velocity, where, typically, different balls have different final velocities. ♢

12.35 Exercise (Billiard Balls)

(a) For n balls of equal mass in one dimension ($d = 1$), give an upper bound for the number of collisions.
Compare with the dynamics of Newton's cradle, the pendulum arrangement shown in the figure on the right.

(b) Same question for three balls of different masses. ♢



In the following, we study the nonrelativistic motion of n particles in $\mathbb{R}^d$ that can attract or repel each other with forces derived from a potential. Again denoting their masses as $m_1, \ldots, m_n > 0$, the Hamiltonian

$$H(p, q) := \sum_{k=1}^{n} \frac{\|p_k\|^2}{2m_k} + V(q) \tag{12.5.2}$$

leads to the equations of motion

$$\dot{p}_k = -\nabla_{q_k} V(q) \quad , \quad \dot{q}_k = \frac{p_k}{m_k} \qquad (k = 1, \ldots, n). \tag{12.5.3}$$

We assume that the potential is composed of two-body potentials, i.e.,

$$V(q) = \sum_{1 \le i < j \le n} V_{i,j}(q_i - q_j)\,.$$

12.36 Example (n-Body Problem)
A prominent example is the *n-body problem of celestial mechanics* from Chapter 11.3.3. There, $V_{i,j}(Q) := -\frac{m_i m_j}{\|Q\|}$, and the Newtonian differential equations are (1.8). In contrast, in *electrostatics*, the particles carry charges $Z_i \in \mathbb{R}$ and $V_{i,j}(Q) := \frac{Z_i Z_j}{\|Q\|}$. So charges of equal sign repel each other, whereas gravitation is always attractive. ♢

The following exercise explains why the gravitational potential of a centrally symmetric mass distribution, for instance (in good approximation) of a star, has the claimed form in the exterior domain.

12.37 Exercise (Potential of a Centrally Symmetric Mass Distribution)
We study the (gravitational or electrostatic) potential generated by a mass or charge distribution. Assume the continuous density $\rho : \mathbb{R}^3 \to [0, \infty)$ to have its support

contained in the ball $B_R^3 = \{x \in \mathbb{R}^3 \mid \|x\| \le R\}$. We assume that this density is centrally symmetric, i.e., $\rho(x) = \tilde{\rho}(\|x\|)$. The potential is given by:

$$V : \mathbb{R}^3 \to \mathbb{R} \quad , \quad V(q) := \int_{B_R^3} \frac{\rho(x)}{\|q - x\|}\, \mathrm{d}x \, ,$$

and we call $M := \int_{B_R^3} \rho(x)\, \mathrm{d}x$ the *mass* or *charge*.

(a) Explain why the potential V is centrally symmetric, i.e., $V(q) = V\big(0, 0, \|q\|\big)$.
(b) We assume $q = (0, 0, a)$, where $a > 0$. Express $\|q - x\|$ in spherical coordinates.
(c) Show by means of spherical coordinates that

$$V(q) = \frac{2\pi}{a} \int_0^R r\tilde{\rho}(r)\big(r + a - |a - r|\big)\, \mathrm{d}r \, .$$

(d) Show that $V(q) = \dfrac{M}{a}$ for $a > R$.
(e) Show for $0 < a \le R$ that $V(q) = \tilde{V}(a)$ where

$$\tilde{V}(a) := 4\pi \left(\int_0^a \frac{r^2}{a} \tilde{\rho}(r)\, \mathrm{d}r + \int_a^R r\tilde{\rho}(r)\, \mathrm{d}r \right) .$$

Derive that $\big(\partial_a^2 \tilde{V}\big)(a) + \frac{2}{a}\big(\partial_a \tilde{V}\big)(a) = -4\pi\tilde{\rho}(a)$ (i.e., in Euclidean coordinates: $-\Delta V = 4\pi\rho$). This is the *Poisson formula*.
(f) Satellites close to the earth have orbit times of about one and a half hours. How long approximately would a satellite orbiting close to the sun take? ◊

The singularities occurring in Example 12.36 complicate the study of the dynamics significantly. Therefore we only consider two-body potentials $V_{i,j}$ that are in $C^2(\mathbb{R}^d, \mathbb{R})$ and assume that they are long range in the sense of Definition 12.1. Thus we require, with appropriate constants $C, \varepsilon > 0$, that

$$|\partial^\alpha V_{i,j}(Q)| \le C \langle Q \rangle^{-|\alpha| - \varepsilon} \qquad \big(1 \le i < j \le n,\ Q \in \mathbb{R}^d,\ \alpha \in \mathbb{N}_0^d,\ |\alpha| \le 2\big) \, .$$

It is convenient to define $V_{j,i}(q) := V_{i,j}(-q)$ when $1 \le i < j \le n$. Then the Hamiltonian (12.5.2) is twice continuously differentiable on the phase space $P := \mathbb{R}_p^{nd} \times \mathbb{R}_q^{nd}$, and by Theorem 11.1, the flow is $\Phi \in C^1(\mathbb{R} \times P, P)$. We write the solutions in the form $\big(p(t, x_0), q(t, x_0)\big) = \Phi(t, x_0)$, or briefly, $\big(p(t), q(t)\big)$.

12.38 Theorem (Motion of Center of Mass) *Denote by*

- $m_N := \sum_{k=1}^n m_k$ *the* **total mass**,
- $p_N : P \to \mathbb{R}^d,\ (p, q) \mapsto \sum_{k=1}^n p_k$ the **total momentum** and
- $q_N : P \to \mathbb{R}^d,\ (p, q) \mapsto \frac{1}{m_N} \sum_{k=1}^n m_k q_k$ the **center of mass**.

Then $\big(p_N(t), q_N(t)\big) := (p_N, q_N) \circ \Phi_t(p,q)$ *satisfies:*

$$p_N(t) = p_N(0) \quad , \quad q_N(t) = q_N(0) + \frac{p_N(0)}{m_N}\, t \qquad (t \in \mathbb{R}).$$

Proof: The equations of motion (12.5.3) imply

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} p_N(t) &= \sum_{k=1}^n \frac{\mathrm{d}}{\mathrm{d}t} p_k(t) = -\sum_{k=1}^n \nabla_{q_k} V(q) = -\sum_{k=1}^n \sum_{1\le i<j\le n} \nabla_{q_k} V_{i,j}(q_i - q_j)\\ &= \sum_{1\le i<j\le n} \big(\nabla_{q_i} V_{i,j}(q_i - q_j) + \nabla_{q_j} V_{i,j}(q_i - q_j)\big) = 0\,.\end{aligned}$$

On the other hand,

$$\frac{\mathrm{d}}{\mathrm{d}t} q_N(t) = \frac{1}{m_N}\sum_{k=1}^n m_k \frac{\mathrm{d}}{\mathrm{d}t} q_k(t) = \frac{1}{m_N}\sum_{k=1}^n p_k(t) = \frac{p_N(0)}{m_N}\,. \qquad \square$$

So the total momentum is a conserved quantity, and we also know the time evolution of the center of mass. As V only depends on the distances $q_i - q_j$, not on q_N, we can reduce the phase space dimension by $2d$ and study the motion in the center of mass system.

12.39 Example (Reduction of the Two Body Problem)
This is done most easily for the motion of two bodies. In terms of the total mass $m_N = m_1 + m_2$ and the *reduced mass* $m_r := \frac{m_1 m_2}{m_1+m_2}$, the linear transformation $\Psi : P \to P$ with phase space $P = T^*\mathbb{R}^{2d}$,
$(p_1, p_2, q_1, q_2) \mapsto (p_N, p_r, q_N, q_r) := \Big(p_1 + p_2, \frac{m_2 p_1 - m_1 p_2}{m_N}, \frac{m_1 q_1 + m_2 q_2}{m_N}, q_1 - q_2\Big)$
leads to the Hamiltonian $H \circ \Psi = H_N + H_r$ with[14]

$$H_N(p_N, q_N) := \frac{\|p_N\|^2}{2m_N} \quad \text{and} \quad H_r(p_r, q_r) := \frac{\|p_r\|^2}{2m_r} + V_{1,2}(q_r)\,.$$

Now $\Psi \in \mathrm{Sp}(P, \omega_0)$, i.e., this linear mapping is symplectic on the phase space P, because $\Psi(p,q) = \tilde\Psi \binom{p}{q}$ with the matrix

$$\tilde\Psi := \begin{pmatrix} A & 0 \\ 0 & D\end{pmatrix} \in \mathrm{Mat}(4d, \mathbb{R}) \quad , \quad A := \begin{pmatrix} \mathbb{1} & \mathbb{1} \\ \frac{m_2}{m_N}\mathbb{1} & -\frac{m_1}{m_N}\mathbb{1}\end{pmatrix} \text{ and } D := \begin{pmatrix} \frac{m_1}{m_N}\mathbb{1} & \frac{m_2}{m_N}\mathbb{1} \\ \mathbb{1} & -\mathbb{1}\end{pmatrix}.$$

Therefore $A^\top D = \mathbb{1}$ and thus $\tilde\Psi^\top \mathbb{J} \tilde\Psi = \mathbb{J}$, see Remark 6.15.2 and Exercise 6.26.b. So the equation of motion remains Hamiltonian even after transformation Ψ, and the study of H is reduced to the study of $H_r : \mathbb{R}^{2d} \to \mathbb{R}$. $\Diamond$

[14] More precisely: $H \circ \Psi = H_N \circ \pi_1 + H_r \circ \pi_2$ with $(\pi_1, \pi_2) : T^*\mathbb{R}^{2d} \to T^*\mathbb{R}^d_{q_N} \times T^*\mathbb{R}^d_{q_r}$.

12.6 * Asymptotic Completeness

> *"Could one not ask whether one of the bodies will always remain in a certain region of the heavens, or if it could just as well travel further and further away forever; whether the distance between two bodies will grow or diminish in the infinite future, or if it instead remains bracketed between certain limits forever? Could one not ask a thousand questions of this kind which would all be solved once one understood how to construct qualitatively the trajectories of the three bodies?"* HENRI POINCARÉ[15]

12.40 Definition *The n-body* problem (12.5.2) *is called* **asymptotically complete**, *if the asymptotic velocities*

$$\overline{v}^{\pm}(x_0) := \lim_{t \to \pm\infty} \frac{q(t, x_0)}{t} \tag{12.6.1}$$

exist for all initial conditions $x_0 \in P$.

12.41 Remarks (Asymptotic Completeness)

1. In the case $n = 2$, asymptotic completeness has been shown already. Namely it follows from Exercise 12.7.1 in conjunction with Example 12.39.
2. The definition of asymptotic completeness is not uniform through the literature. Our definition is a rather weak one.
3. In quantum mechanical scattering theory, the proof of an analogous property of asymptotic completeness was achieved first, with contributions by V. Enss, Ch. Gérard, G.-M. Graf, I.M. Sigal, A. Soffer, D. Yafaev and others. In the classical case, essential contributions to the proof were provided among others by J. Dereziński and W. HUNZIKER [Hun].
4. The following proof is a variant of the exposition in the standard reference [DG] by DEREZIŃSKI and GÉRARD. It is complicated and may be skipped without problems, since later chapters will not refer to it any more. On the other hand, the entire scattering theory in practice relies on the assumption of asymptotic completeness.

   This proof is among the *highlights* of mathematical physics, and readers can test their analytical skills by trying to simplify it. ◊

In Example 12.39, for two bodies, internal and external dynamics (with Hamiltonians H_r and H_N respectively) have been studied separately. This approach is now to be generalized to the dynamics within and between clusters. So we will do some kinematic deliberations.

---

[15] After: Henri Poincaré: New Methods of Celestial Mechanics, Daniel L. Goroff, Ed., American Institute of Physics. Page 19.

12.42 Definition

- A **partition** or **cluster decomposition** of the index set $N := \{1, \dots, n\}$ is a set $\mathcal{C} := \{C_1, \dots, C_k\}$ of **atoms (clusters)** $\emptyset \neq C_\ell \subseteq N$ such that

$$\bigcup_{\ell=1}^{k} C_\ell = N \quad \text{and} \quad C_\ell \cap C_m = \emptyset \quad \text{for} \quad \ell \neq m\,.$$

- The **lattice of partitions** $\mathcal{P}(N)$ is the set of cluster decompositions $\mathcal{C}$ of N, partially ordered by **refinement**, i.e.,

$$\mathcal{C} = \{C_1, \dots, C_k\} \preccurlyeq \{D_1, \dots, D_\ell\} = \mathcal{D}\,,$$

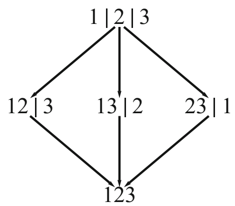
  if $C_m \subseteq D_{\pi(m)}$ for an appropriate mapping $\pi : \{1, \dots, k\} \to \{1, \dots, \ell\}$. In this case, $\mathcal{C}$ is called **finer** than $\mathcal{D}$ and $\mathcal{D}$ **coarser** than $\mathcal{C}$.
- The **rank** of $\mathcal{C} \in \mathcal{P}(N)$ is the number $|\mathcal{C}|$ of its atoms.
- The **join** of $\mathcal{C}$ and $\mathcal{D} \in \mathcal{P}(N)$, denoted as $\mathcal{C} \vee \mathcal{D}$, is the finest cluster decomposition that is coarser than both $\mathcal{C}$ and $\mathcal{D}$.

The unique finest and coarsest elements of $\mathcal{P}(N)$ are

$$\mathcal{C}_{\min} := \big\{\{1\}, \dots, \{n\}\big\} \quad \text{and} \quad \mathcal{C}_{\max} := \big\{\{1, \dots, n\}\big\}$$

respectively.

12.43 Example (Lattice of Partitions)
For $n = 3$, $\mathcal{P}(N)$ consists of five elements, which are ordered as in the graphics to the right (vertically arranged by rank).
In the figure we have for instance used 12|3 to denote the partition $\{\{1, 2\}, \{3\}\}$, with rank $\big|\{\{1, 2\}, \{3\}\}\big| = 2$.
The join of 12|3 and 1|23 = 23|1 is 123. ◊



On the configuration space $M := \bigoplus_{k=1}^{n} M_k = \mathbb{R}^{nd}$ of all particles, where $M_k := \mathbb{R}^d$ is the configuration space of the k^{th} particle, we now want to make a choice of coordinates that is adapted to the cluster decomposition $\mathcal{C} \in \mathcal{P}(N)$, in analogy to Example 12.39. We write configurations $q \in M$ in the form $q = (q_1, \dots, q_n)$ with $q_k \in M_k$. The notation will be simplified by using the mass-weighted scalar product

$$\langle q, q' \rangle_{\mathcal{M}} := \sum_{k=1}^{n} m_k \langle q_k, q'_k \rangle \qquad (q, q' \in M) \tag{12.6.2}$$

and correspondingly the norm $\|q\|_{\mathcal{M}} := \sqrt{\langle q, q \rangle_{\mathcal{M}}}$.

12.44 Definition *In the cluster decomposition* $\mathcal{C} = \{C_1, \ldots, C_k\} \in \mathcal{P}(N)$,

- the **mass** of the i^{th} **cluster** C_i is $m_{C_i} := \sum_{j \in C_i} m_j$.
- *the* **center of mass** of C_i *is* $q^E_{C_i} := \frac{1}{m_{C_i}} \sum_{j \in C_i} m_j q_j$,
- and the **center of mass projection** is the linear mapping

$$\Pi^E_{\mathcal{C}} : M \to M \quad , \quad \Pi^E_{\mathcal{C}}(q)_\ell := q^E_{C_i} \quad , \text{with} \quad \ell \in C_i .$$

Thus $\Pi^E_{\mathcal{C}}$ is an orthogonal projection with respect to the chosen scalar product, and so is $\Pi^I_{\mathcal{C}} := \mathbb{1}_M - \Pi^E_{\mathcal{C}}$. In terms of the latter, $\Pi^I_{\mathcal{C}}(q)_\ell = q_\ell - q^E_{C_i}$ (with $\ell \in C_i$) is the distance of the ℓ^{th} particle from the center of mass of its cluster.

The symbols I and E stand for the *i*nternal and *e*xternal dynamics of clusters.

12.45 Exercise (Cluster Projections) Prove that for all $\mathcal{C} \in \mathcal{P}(N)$, the linear mappings $\Pi^I_{\mathcal{C}}$ and $\Pi^E_{\mathcal{C}}$ are orthogonal projections. ♢

We denote the images of these projections as $\Delta^{(0)}_{\mathcal{C}} := \Pi^E_{\mathcal{C}}(M)$ and call them *collision subspaces*. As they are parametrized by giving the centers of mass of the clusters (of which there are $|\mathcal{C}|$), one has $\dim\big(\Delta^{(0)}_{\mathcal{C}}\big) = d|\mathcal{C}|$, and

$$\Delta^{(0)}_{\mathcal{C}} \cap \Delta^{(0)}_{\mathcal{D}} = \Delta^{(0)}_{\mathcal{C}\vee\mathcal{D}} \qquad \big(\mathcal{C}, \mathcal{D} \in \mathcal{P}(N)\big). \tag{12.6.3}$$

The latter identity is the reason why the $\vee$-operation is relevant, see Figure 12.6.1.

12.46 Remarks

1. By (12.6.3), the subspaces $\Delta^{(0)}_{\mathcal{C}}$ of M generate a partition of M with atoms

$$\Xi^{(0)}_{\mathcal{C}} := \Delta^{(0)}_{\mathcal{C}} \Big\backslash \bigcup_{\mathcal{D} \supsetneq \mathcal{C}} \Delta^{(0)}_{\mathcal{D}} \qquad \big(\mathcal{C} \in \mathcal{P}(N)\big). \tag{12.6.4}$$

   If the asymptotic velocities in (12.6.1) exist, we can partition the particles in the far future or past uniquely into clusters by

$$\mathcal{C}^\pm : P \to \mathcal{P}(N) \quad , \quad \overline{v}^\pm(x_0) \in \Xi^{(0)}_{\mathcal{C}^\pm(x_0)} .$$

   In this asymptotics, there will be no more interaction between particles of different clusters.
2. For *finite times*, the partition (12.6.4) is not very useful for proving anything. Accordingly, we introduce for given $\varepsilon > 0$ the ε-neighborhoods

$$\Delta^{(\varepsilon)}_{\mathcal{C}} := \big\{q \in M \mid \|\Pi^I_{\mathcal{C}}(q)\|_{\mathcal{M}} \le \varepsilon\big\} \qquad \big(\mathcal{C} \in \mathcal{P}(N)\big) . \tag{12.6.5}$$

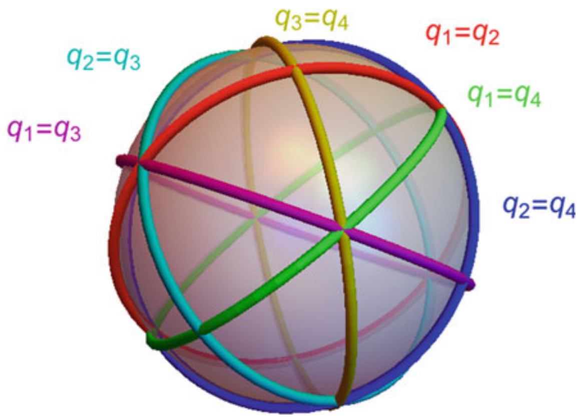


Figure 12.6.1 Collision subspaces for $n = 4$ particles in $d = 1$ dimension, with center of mass at $0 \in \mathbb{R}^4$. Shown are the intersections of these hyperplanes $\{q_i = q_j\}$ with the sphere S^2

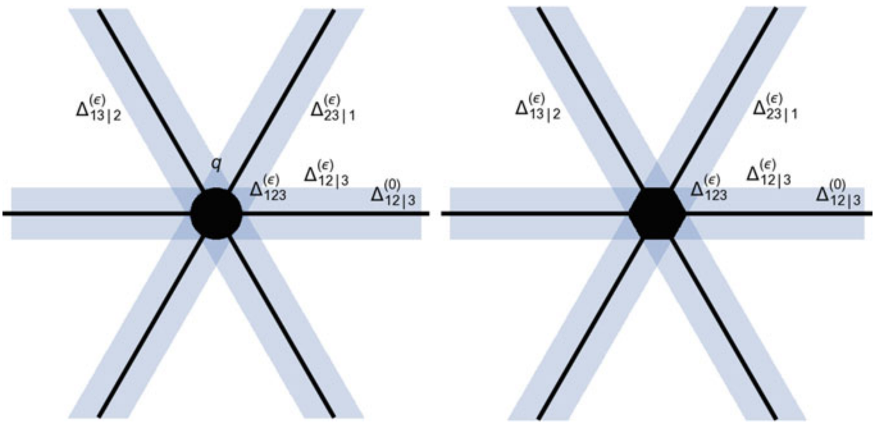


Figure 12.6.2 Inconvenient cluster partitions of the configuration space (with center of mass at 0) for $n = 3$ particles in $d = 1$ dimension

So for $q \in \Delta^{(\varepsilon)}_{\mathcal{C}}$, the position q_i of the i^{th} particle is not too far away from the center of mass of its $\mathcal{C}$-cluster; see Figure 12.6.2 (left).

In the quest for a cluster partition of M, we might be tempted to use, analogous to (12.6.4), the subsets

$$\Xi^{(\varepsilon)}_{\mathcal{C}} \overset{?}{:=} \Delta^{(\varepsilon)}_{\mathcal{C}} \Big\backslash \bigcup_{\mathcal{D} \supsetneq \mathcal{C}} \Delta^{(\varepsilon)}_{\mathcal{D}} \qquad \big(\mathcal{C} \in \mathcal{P}(N)\big)\,. \tag{12.6.6}$$

However, these are not disjoint, because the analog of (12.6.3) does not hold for the $\Delta_{\mathcal{C}}^{(\varepsilon)}$. Figure 12.6.2 (left) shows that there are points $q \in M$ that cannot be uniquely assigned to one $\Xi_{\mathcal{C}}^{(\varepsilon)}$, and that the set of these points even has positive volume.

For the same reason, it is not helpful either to modify Definition (12.6.5), beginning with the family of ε-neighborhoods $\Delta_{\mathcal{D}}^{(\varepsilon)}$ of *two*-particle collision subspaces $\Delta_{\mathcal{D}}^{(0)}$, $|\mathcal{D}| = n-1$, by defining

$$\Delta_{\mathcal{C}_{\min}}^{(\varepsilon)} \overset{?}{:=} M \quad \text{and} \quad \Delta_{\mathcal{C}}^{(\varepsilon)} \overset{?}{:=} \bigcap_{\mathcal{D} \preccurlyeq \mathcal{C},\ |\mathcal{D}|=n-1} \Delta_{\mathcal{D}}^{(\varepsilon)} \qquad \big(\mathcal{C} \in \mathcal{P}(N), |\mathcal{C}| < n-1\big),$$

see Figure 12.6.2 (right). Applied to this family, (12.6.6) again does not yield a partition of M. ♢

This is why we proceed differently and use the *Graf partition* of M: This partition relies on the *(mean) moment of inertia*

$$J : M \to \mathbb{R} \quad , \quad J(q) = \|q\|_{\mathcal{M}}^2 = \sum_{k=1}^{n} m_k \|q_k\|^2 \, ,$$

which, in the cluster decomposition $\mathcal{C}$, will be of the form

$$J = J_{\mathcal{C}}^E + J_{\mathcal{C}}^I \quad \text{with} \quad J_{\mathcal{C}}^E := J \circ \Pi_{\mathcal{C}}^E \quad \text{and} \quad J_{\mathcal{C}}^I := J \circ \Pi_{\mathcal{C}}^I \, , \tag{12.6.7}$$

due to the orthogonality of the projection $\Pi_{\mathcal{C}}^E$. Indeed,

$$J(q) = \big\langle (\Pi_{\mathcal{C}}^E + \Pi_{\mathcal{C}}^I) q \, , (\Pi_{\mathcal{C}}^E + \Pi_{\mathcal{C}}^I) q \big\rangle_{\mathcal{M}} \, ,$$

and

$$\big\langle \Pi_{\mathcal{C}}^E q \, , \Pi_{\mathcal{C}}^I q \big\rangle_{\mathcal{M}} = \big\langle \Pi_{\mathcal{C}}^E q \, , (\mathbb{1}_{\mathcal{M}} - \Pi_{\mathcal{C}}^E) q \big\rangle_{\mathcal{M}} = \big\langle q \, , \Pi_{\mathcal{C}}^E (\mathbb{1}_{\mathcal{M}} - \Pi_{\mathcal{C}}^E) q \big\rangle_{\mathcal{M}} = 0 \, .$$

Here we have the intuitive understanding

- of $J_{\mathcal{C}}^E(q)$ as the moment of inertia of the configuration in which all masses of each cluster are joined in its center of mass;
- of $J_{\mathcal{C}}^I(q)$ as the sum of the moments of inertia of the clusters, each referred to the respective center of mass, rather than the origin.

12.47 Exercise (Moments of Inertia) Show that for partitions $\mathcal{C}, \mathcal{D} \in \mathcal{P}(N)$ with $D \succcurlyeq \mathcal{C}$, one has $J_{\mathcal{D}}^E \le J_{\mathcal{C}}^E$ (and hence by (12.6.7) $J_{\mathcal{D}}^I \ge J_{\mathcal{C}}^I$). ♢

12.48 Lemma *There exists* $\delta' \in (0, \frac{1}{2}]$ *such that*

$$\delta'\big(J^I_{\mathcal{C}} + J^I_{\mathcal{D}}\big) \le J^I_{\mathcal{C}\vee\mathcal{D}} \le \frac{J^I_{\mathcal{C}} + J^I_{\mathcal{D}}}{\delta'} \qquad \big(\mathcal{C}, \mathcal{D} \in \mathcal{P}(N)\big).$$

Proof:
• By Exercise 12.47, any constant $\delta' \in (0, \frac{1}{2}]$ can be used for the left inequality.
• For the right inequality, we use the identities $J^I_{\mathcal{C}\vee\mathcal{D}} = J \circ \Pi^I_{\mathcal{C}\vee\mathcal{D}}$ and $(J^I_{\mathcal{C}} + J^I_{\mathcal{D}}) \circ \Pi^I_{\mathcal{C}\vee\mathcal{D}} = J^I_{\mathcal{C}} + J^I_{\mathcal{D}}$. Both $J^I_{\mathcal{C}\vee\mathcal{D}}$ and $J^I_{\mathcal{C}} + J^I_{\mathcal{D}}$ are positive definite on the space $\mathrm{im}(\Pi^I_{\mathcal{C}\vee\mathcal{D}}) = \ker(\Pi^E_{\mathcal{C}\vee\mathcal{D}})$.
Such quadratic forms on finite dimensional vector spaces are always comparable. So for every pair $(\mathcal{C}, \mathcal{D})$, there exists a constant $\delta'(\mathcal{C}, \mathcal{D}) \in (0, \frac{1}{2}]$ that makes the right inequality true. But there are only finitely many pairs, so there exists a smallest nonzero δ'. □

12.49 Definition For $\delta \in (0, 1)$, let

$$J^{(\delta)} : M \to \mathbb{R} \quad , \quad J^{(\delta)}(q) := \max\big\{J^E_{\mathcal{C}}(q) + \delta^{|\mathcal{C}|} \bigm| \mathcal{C} \in \mathcal{P}(N)\big\}.$$

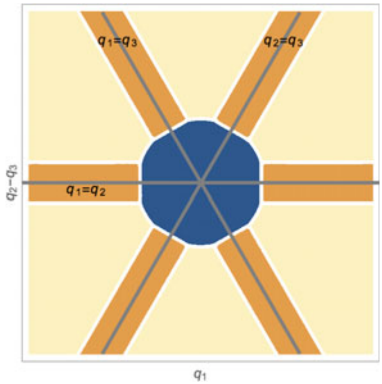
The **Graf partition** of the configuration space M is the family of subsets

$$\Xi^{(\delta)}_{\mathcal{C}} := \Big\{q \in M \Bigm| J^E_{\mathcal{C}}(q) + \delta^{|\mathcal{C}|} = J^{(\delta)}(q)\Big\} \qquad \big(\mathcal{C} \in \mathcal{P}(N)\big). \tag{12.6.8}$$

While this is not a partition in the sense of sets, it is a partition in the sense of measure theory since

$$\bigcup_{\mathcal{C}\in\mathcal{P}(N)} \Xi^{(\delta)}_{\mathcal{C}} = M\,,$$

and for $\mathcal{C} \neq \mathcal{D}$ the Lebesgue measure of $\Xi^{(\delta)}_{\mathcal{C}} \cap \Xi^{(\delta)}_{\mathcal{D}}$ is zero, because values of the functions $J^E_{\mathcal{C}} + \delta^{|\mathcal{C}|}$ and $J^E_{\mathcal{D}} + \delta^{|\mathcal{D}|}$ coincide only on quadrics in M.
We expect from a cluster decomposition of M that particles in the same cluster are close to each other, whereas particles of different clusters stay apart by some minimum distance.
Both expectations are met, see the figure on the right and the following lemma.



Graf partition of the configuration space (center of mass at 0) for $n = 3$ particles in $d = 1$ dimension.
Yellow: $\Xi^{(\delta)}_{\mathcal{C}_{\min}}$, Blue: $\Xi^{(\delta)}_{\mathcal{C}_{\max}}$

12.50 Lemma *For every* $\delta \in \left(0, \frac{\delta'}{4}\right)$, *there exist* $0 < \rho_1 < \rho_2$ *such that the* ε-*neighborhoods* $\Delta_{\mathcal{C}}^{(\varepsilon)}$ *of the collision subspaces* $\Delta_{\mathcal{C}}^{(0)}$ *as defined in (12.6.5) satisfy (see Figure 12.6.3)*

$$\Delta_{\mathcal{C}}^{(\rho_1)} \subseteq \bigcup_{\mathcal{D} \succcurlyeq \mathcal{C}} \Xi_{\mathcal{D}}^{(\delta)} \subseteq \Delta_{\mathcal{C}}^{(\rho_2)} \qquad \big(\mathcal{C} \in \mathcal{P}(N)\big).$$

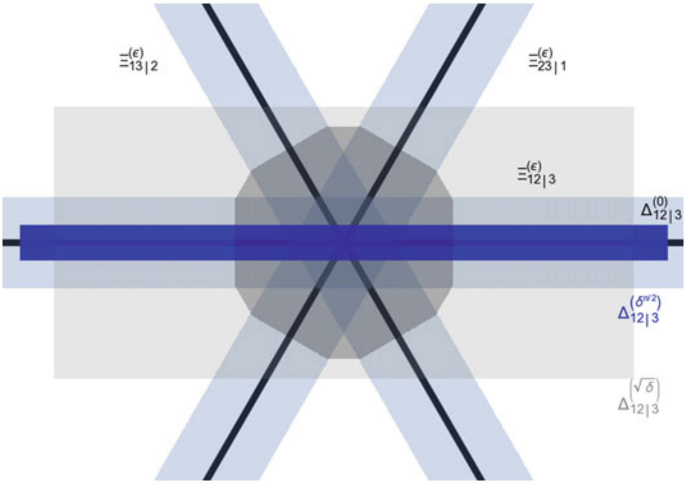


Figure 12.6.3 The inclusions of Lemma 12.50 for $n = 3$, $d = 1$ and $\mathcal{C} = \{\{1, 2\}, \{3\}\}$

12.51 Remark Readers should convince themselves that the lemma shows that the Graf partition indeed satisfies the above requirements for a cluster decomposition. ◊

Proof:

• We begin with the inclusion on the right and set $\rho_2 := \sqrt{\delta}$. Then $\Delta_{\mathcal{C}}^{(\rho_2)} = \{q \in M \mid J_{\mathcal{C}}^I(q) \le \delta\}$. If $\mathcal{D} \succcurlyeq \mathcal{C}$, then by Exercise 12.47, one has $J_{\mathcal{D}}^I(q) \ge J_{\mathcal{C}}^I(q)$. But for $q \in \Xi_{\mathcal{D}}^{(\delta)}$, one has (by definition of the Graf partition and comparing $\mathcal{D}$ with $\mathcal{C}_{\min}$) the estimate $J_{\mathcal{D}}^I(q) \le \delta^{|\mathcal{D}|} - \delta^n \le \delta$. Combining the two, we obtain $J_{\mathcal{C}}^I(q) \le \delta$, hence $q \in \Delta_{\mathcal{C}}^{(\rho_2)}$.

• Instead of the left inclusion, we show this slightly stronger (stronger since the partition is not entirely disjoint) statement: For an appropriate $\rho_1 \in (0, \rho_2)$, one has

$$\Delta_{\mathcal{C}}^{(\rho_1)} \cap \Xi_{\mathcal{E}}^{(\delta)} = \emptyset \qquad \text{if } \mathcal{E} \not\succcurlyeq \mathcal{C} \quad (\mathcal{E} \text{ not coarser than } \mathcal{C})\,. \tag{12.6.9}$$

Under this condition, one always has $|\mathcal{E} \vee \mathcal{C}| < |\mathcal{E}|$ and $|\mathcal{E} \vee \mathcal{C}| \le |\mathcal{C}|$, with equality if and only if $\mathcal{E} \precneqq \mathcal{C}$ ($\mathcal{E}$ strictly finer than $\mathcal{C}$).

For $q \in \Delta_{\mathcal{C}}^{(\rho_1)} \cap \Xi_{\mathcal{E}}^{(\delta)}$, one has $J_{\mathcal{C}}^I(q) \le \rho_1^2$ and, by definition of $\Xi_{\mathcal{E}}^{(\delta)}$,

$$J_{\mathcal{E}}^I(q) \le J_{\mathcal{E}\vee\mathcal{C}}^I(q) - \delta^{|\mathcal{E}\vee\mathcal{C}|}\left(1 - \delta^{|\mathcal{E}|-|\mathcal{E}\vee\mathcal{C}|}\right). \tag{12.6.10}$$

• Now if we had $\mathcal{E} \precneqq \mathcal{C}$, then the condition in (12.6.9) would be satisfied, $\mathcal{E} \vee \mathcal{C} = \mathcal{C}$, and the inequality would imply

$$J_E^I(q) \le J_{\mathcal{C}}^I(q) - \delta^{|\mathcal{C}|}(1-\delta) \le \rho_1^2 - \delta^{|\mathcal{C}|}(1-\delta) \le \rho_1^2 - \delta^{n-1}(1-\delta) < 0$$

for small ρ_1. But this is not possible because $J_E^I \ge 0$.

• The other case that is compatible with condition (12.6.9), namely that $\mathcal{C}$ and $\mathcal{E}$ are not comparable and one has $\mathcal{E} \not\preccurlyeq \mathcal{C}$ in addition to $\mathcal{E} \not\succcurlyeq \mathcal{C}$, also leads to a contradiction for small ρ_1, because with Lemma 12.48, one concludes from (12.6.10) for $\delta < \frac{1}{4}\delta' \le \frac{1}{8}$ that

$$\begin{aligned}
J_{\mathcal{E}}(q) &\le \frac{1}{\delta'}\left(J_{\mathcal{E}}^I(q) + J_{\mathcal{C}}^I(q)\right) - \delta^{|\mathcal{E}\vee\mathcal{C}|}(1-\delta) \\
&\le \frac{1}{\delta'}\left(2J_{\mathcal{C}}^I(q) + \delta^{|\mathcal{E}|} - \delta^{|\mathcal{C}|}\right) - \delta^{|\mathcal{E}\vee\mathcal{C}|}(1-\delta) \le \frac{2}{\delta'}\rho_1^2 + \frac{\delta^{|\mathcal{E}|}}{\delta'} - \tfrac{1}{2}\delta^{|\mathcal{E}\vee\mathcal{C}|} \\
&\le \frac{2}{\delta'}\rho_1^2 - \tfrac{1}{4}\delta^{|\mathcal{E}\vee\mathcal{C}|} \le \frac{2}{\delta'}\rho_1^2 - \tfrac{1}{4}\delta^{n-1}\,.
\end{aligned}$$

In the second last inequality we have used $|\mathcal{E}| \ge |\mathcal{E} \vee \mathcal{C}|$. This expression will become negative, too, if we choose $\rho_1 < \sqrt{\delta^n \delta'}$. So we have led the assumption $q \in \Delta_{\mathcal{C}}^{(\rho_1)} \cap \Xi_{\mathcal{E}}^{(\delta)}$ to a contradiction. □

Another property that was illustrated in the figure on page 316 is generally valid for Graf partitions:

12.52 Lemma *For small* $\delta \in (0,1)$*, the Graf partition (12.6.8) has the property that for* $\Xi_{\mathcal{C}}^{(\delta)} \cap \Xi_{\mathcal{D}}^{(\delta)} \ne \emptyset$*, the cluster decompositions* $\mathcal{C}$ *and* $\mathcal{D}$ *are* **comparable**, i.e., $\mathcal{C} \preccurlyeq \mathcal{D}$ or $\mathcal{C} \succcurlyeq \mathcal{D}$.

Proof: By Lemma 12.48, for $q \in \Xi_{\mathcal{C}}^{(\delta)} \cap \Xi_{\mathcal{D}}^{(\delta)}$, we have

$$J_{\mathcal{C}\vee\mathcal{D}}^I(q) \le \frac{J_{\mathcal{C}}^I(q) + J_{\mathcal{D}}^I(q)}{\delta'} \quad \text{and} \quad J_{\mathcal{D}}^I(q) - \delta^{|\mathcal{D}|} = J_{\mathcal{C}}^I(q) - \delta^{|\mathcal{C}|}\,.$$

Now if $\mathcal{C}$ and $\mathcal{D}$ were not comparable, and therefore $|\mathcal{C} \vee \mathcal{D}| < \min(|\mathcal{C}|, |\mathcal{D}|)$, and without loss of generality $|\mathcal{C}| \le |\mathcal{D}|$, then one would conclude:

$$\begin{aligned}
&\left(J_{\mathcal{C}\vee\mathcal{D}}^I(q) - \delta^{|\mathcal{C}\vee\mathcal{D}|}\right) - \left(J_{\mathcal{C}}^I(q) - \delta^{|\mathcal{C}|}\right) \\
&\quad\le \frac{1}{\delta'}\left(J_{\mathcal{C}}^I(q) + \left(J_{\mathcal{C}}^I(q) - \delta^{|\mathcal{C}|} + \delta^{|\mathcal{D}|}\right)\right) - \left(J_{\mathcal{C}}^I(q) - \delta^{|\mathcal{C}|}\right) - \delta^{|\mathcal{C}\vee\mathcal{D}|} \\
&\quad\le \left(\frac{2}{\delta'} - 1\right) J_{\mathcal{C}}^I(q) + \delta^{|\mathcal{C}|} - \delta^{|\mathcal{C}\vee\mathcal{D}|} \le \left(\frac{2}{\delta'} - 1\right)\delta^{|\mathcal{C}|} + \delta^{|\mathcal{C}|} - \delta^{|\mathcal{C}\vee\mathcal{D}|} \\
&\quad\le \left(\frac{2\delta}{\delta'} - 1\right)\delta^{|\mathcal{C}\vee\mathcal{D}|} < 0\,,
\end{aligned}$$

provided $\delta \in (0, \delta'/2)$. This however means that q cannot lie in $\Xi_{\mathcal{C}}^{(\delta)}$, because in Definition 12.49, $J_{\mathcal{C}}^{E}(q) + \delta^{|\mathcal{C}|} = J(q) - \big(J_{\mathcal{C}}^{I}(q) - \delta^{|\mathcal{C}|}\big)$ is not maximal. □

It is because of this lemma that the Graf partition of the configuration space M is closely related to the lattice of partitions $\mathcal{P}(N)$. We will use the Graf partition in the proof of the following main statement:

12.53 Theorem (Asymptotic Completeness) *For all Hamiltonians (12.5.2) with long range potentials $V_{i,j} \in C^2(\mathbb{R}^d, \mathbb{R})$ $(1 \le i < j \le n)$ and for all initial conditions $x_0 \in P$, there exist asymptotic velocities $\overline{v}^{\pm}(x_0) = \lim_{t\to\pm\infty} \frac{q(t,x_0)}{t}$.*

Proof:

• As the Hamiltonian flow $\Phi \in C^2(\mathbb{R} \times P, P)$ is reversible, it suffices to show the existence of $\overline{v}^{+}(x_0)$.

• In a first step, we show that the *norm* of $t \mapsto \frac{q(t,x_0)}{t}$ has a limit. With $\big(p(t), q(t)\big) := \Phi(t, x_0)$, define

$$j(t) := J\left(\frac{q(t)}{t}\right) \qquad \big(t \in [1, \infty)\big).$$

So we want to show the existence of $j^+ := \lim_{t\to\infty} j(t)$.

• The key step is a cluster decomposition that depends on time and is adjusted to the motion. We will find measurable mappings

$$\mathcal{A} : [1, \infty) \to \mathcal{P}(N) \quad \text{with} \quad \frac{q(t)}{t^{1-\varepsilon/2}} \in \Xi_{\mathcal{A}(t)}^{(\delta)} . \tag{12.6.11}$$

It is exactly on the interfaces between the $\Xi_{\mathcal{C}}^{(\delta)}$ that $\mathcal{A}$ is not uniquely determined by this condition. But since the trajectory q in position space is twice continuously differentiable (and since, by reason of Lemma 12.52, $\delta^{|\mathcal{C}|} \ne \delta^{|\mathcal{D}|}$ when $\mathcal{C} \ne \mathcal{D}$ and $\Xi_{\mathcal{C}}^{(\delta)} \cap \Xi_{\mathcal{D}}^{(\delta)} \ne \emptyset$ in (12.6.8)), we can conclude by Sard's theorem [16] that for almost all $\delta \in (0, 1)$ in Definition 12.49 of the Graf partition, $\mathcal{A}$ can even be chosen to be piecewise constant.

• The scaling factor $t^{1-\varepsilon/2}$ in (12.6.11) has the consequence that in the decomposition of the moment of inertia,

$$j(t) = j^E(t) + j^I(t) \quad \text{with} \quad j^E(t) := J_{\mathcal{A}(t)}^{E}\left(\tfrac{q(t)}{t}\right) \text{ and } j^I(t) := J_{\mathcal{A}(t)}^{I}\left(\tfrac{q(t)}{t}\right), \tag{12.6.12}$$

the second term does not contribute to the asymptotics:

$$j^I(t) = t^{-\varepsilon} J_{\mathcal{A}(t)}^{I}\left(\tfrac{q(t)}{t^{1-\varepsilon/2}}\right) \le t^{-\varepsilon} , \tag{12.6.13}$$

[16] **Theorem (Sard)**, see Hirsch [Hirs, Chapter 3.1]: *For $f \in C^k(U, \mathbb{R}^m)$ with $k \ge \max(n-m+1, 1)$ and $U \subseteq \mathbb{R}^n$ open, let* $\mathrm{Crit}(f) := \{x \in U \mid \mathrm{rank}(\mathrm{D}f(x)) < m\}$ *be the* **critical set** *of f. Then the set $f\big(\mathrm{Crit}(f)\big) \subseteq \mathbb{R}^m$ of* **critical values** *has Lebesgue measure* 0..

because for all $\mathcal{C} \in \mathcal{P}(N)$ and $\tilde{q} \in \Xi_{\mathcal{C}}^{(\delta)}$, it follows by (12.6.7) and the definition of the Graf partition that

$$J_{\mathcal{C}}^{I}(\tilde{q}) = J(\tilde{q}) - J_{\mathcal{C}}^{E}(\tilde{q}) = J_{\mathcal{C}_{\min}}^{E}(\tilde{q}) - J_{\mathcal{C}}^{E}(\tilde{q}) \leq \delta^{|\mathcal{C}|} - \delta^{n} \leq 1\,. \tag{12.6.14}$$

To show the existence of the limit $j^{+} = \lim_{t\to\infty} j(t)$, it therefore suffices, by (12.6.12), to show that $j^{E}(t)$ has a limit. In doing so, one has the difficulty that j^{E} is discontinuous where $t \mapsto \mathcal{A}(t)$ is not constant.

We therefore decompose j^{I} into

$$j^{I} = \tilde{j}^{I} + h \tag{12.6.15}$$

with $\tilde{j}^{I}$ continuous and h piecewise constant. This decomposition is unique up to an additive constant, which we will yet determine. Accordingly, in view of $j^{E} = j - j^{I}$, the function $\tilde{j}^{E} := j^{E} - h$ is continuous as well.

With $q^{I}(t) := \Pi^{I}_{\mathcal{A}(t)} q(t)$ and $v^{I}(t) := \frac{\mathrm{d}}{\mathrm{d}t} q^{I}(t)$, we get the piecewise equality

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{j}^{I}(t) = \frac{2}{t}\left\langle \frac{q^{I}(t)}{t}, v^{I}(t) - \frac{q^{I}(t)}{t}\right\rangle_{\mathcal{M}} = \mathcal{O}\left(t^{-1-\varepsilon/2}\right)\,, \tag{12.6.16}$$

because $q^{I}(t) = \mathcal{O}(t^{1-\varepsilon/2})$ and

$$\|v^{I}(t)\|_{\mathcal{M}} \leq \|v(t)\|_{\mathcal{M}} = \sqrt{2\big(E - V(q(t))\big)} \leq \sqrt{2(E - V_{\min})}\,.$$

Therefore, $\lim_{t\to\infty} \tilde{j}^{I}(t)$ exists as well, and we choose the additive constant in (12.6.15) in such a way as to make this limit 0. Then we have also

$$h(t) = j^{I}(t) - \tilde{j}^{I}(t) = \mathcal{O}(t^{-\varepsilon}) + \mathcal{O}(t^{-\varepsilon/2}) = \mathcal{O}(t^{-\varepsilon/2})\,. \tag{12.6.17}$$

• To show that the limit of j^{E} exists, it suffices to show that the continuous function $t \mapsto \tilde{j}^{E}(t) = \tilde{j}^{E}(1) + \int_1^t f(s)\,\mathrm{d}s$ with

$$f(s) := \frac{2}{s}\left\langle \frac{q^{E}(s)}{s}, v^{E}(s) - \frac{q^{E}(s)}{s}\right\rangle_{\mathcal{M}} = \mathcal{O}(1/s) \tag{12.6.18}$$

(and $q^{E} := q - q^{I}$, $v^{E} := \frac{\mathrm{d}}{\mathrm{d}t} q^{E}$) converges. We again decompose the integrand as $f = \tilde{f} + g$ with $\tilde{f}$ continuous and g piecewise constant. Then we get

$$f(t) = f(1) + g(t) - g(1) + \int_1^t I(s)\,\mathrm{d}s \tag{12.6.19}$$

with the integrand $I(s) := I_1(s) + I_2(s) + I_3(s) = \mathcal{O}(1/s^2)$, $I_1(s) := -\frac{2}{s} f(s)$,

$$I_2(s) := \frac{2}{s^2}\left\| v^E(s) - \frac{q^E(s)}{s}\right\|^2_{\mathcal{M}} \quad , \quad I_3(s) := -\frac{2}{s}\left\langle \frac{q^E(s)}{s}, \nabla V^E(q(s))\right\rangle,$$

and the *intercluster potential*

$$V^E(q(s)) := \sum\nolimits'_{1\le i<j\le n} V_{i,j}\,(q_i(s) - q_j(s))\,.$$

The symbol $\sum'$ indicates that we only sum over those pairs (i, j) of indices that belong to different atoms in the cluster decomposition $\mathcal{A}(s)$. For these pairs, one has by (12.6.11) and Lemma 12.50 the estimates

$$\|q_i(s) - q_j(s)\| \ge c\, s^{1-\varepsilon/2} \qquad \big(s \in [1,\infty)\big),$$

hence by the long range property (Definition 12.1) of the pair potentials,

$$\nabla V^E\big(q(s)\big) = \mathcal{O}\big((s^{1-\varepsilon/2})^{-1-\varepsilon}\big) = \mathcal{O}\big(s^{-1-\frac{\varepsilon}{4}}\big)\,, \tag{12.6.20}$$

provided we choose $\varepsilon \in (0, \frac{1}{2})$.

• The integrand I in (12.6.19) is piecewise equal to the time derivative of f. As can be checked by plugging in, the integral equation (12.6.19) has the solution

$$f(t) = g(t) + t^{-2}\left(f(1) - g(1) + \int_1^t \big(s^2\,(I_2(s) + I_3(s)) - 2s\, g(s)\big)\,\mathrm{d}s\right). \tag{12.6.21}$$

We would like to show that this solution satisfies $f \in L^1\big([1,\infty)\big)$, because then $t \mapsto \tilde{j}^E(t)$ converges.

For $g \equiv 0$, we argue as follows:

- The absolute value of I_3, being integrable, see (12.6.20), leads to an integrable contribution of order $\mathcal{O}(t^{-1-\frac{\varepsilon}{4}})$ in (12.6.21).
- I_2 is non-negative. If the contribution from I_2 were to cause f to be non-integrable, then $\tilde{j}^E$ would diverge to $+\infty$. But we know that $\tilde{j}^E$ is bounded. Therefore $\tilde{j}^E$ converges.

• If the parameter ε quantifying the long range property were 0, then the jump of g at a point of discontinuity t_0 would satisfy:

$$g(t_0^+) - g(t_0^-) = g_A(t_0^+) - g_A(t_0^-) \qquad \text{with } g(t_0^\pm) := \lim_{\gamma\searrow 0} g(t_0 \pm \gamma)\ \text{ etc.}$$

and $g_A(t_0^\pm) := t_0^{-\varepsilon}\frac{\mathrm{d}}{\mathrm{d}t} J^E_{\mathcal{A}(t_0^\pm)}\left(\frac{q(t)}{t}\right)\Big|_{t=t_0}$.

In reality however, $\varepsilon > 0$ and therefore by (12.6.18),

$$g(t_0^+) - g(t_0^-) = g_A(t_0^+) - g_A(t_0^-)\ +\ g_B(t_0^+) - g_B(t_0^-) \tag{12.6.22}$$

for the piecewise constant function g_B with the jumps

$$g_B(t_0^+) - g_B(t_0^-) = -\frac{\varepsilon}{t_0}\left[\left\langle\frac{q^E(t_0^+)}{t_0}, v^E(t_0^+)\right\rangle_{\mathcal{M}} - \left\langle\frac{q^E(t_0^-)}{t_0}, v^E(t_0^-)\right\rangle_{\mathcal{M}}\right].$$

– With the h from (12.6.15), one has

$$g_B(t_2) - g_B(t_1) = -2\varepsilon\left[\frac{h(t_2)}{t_2} - \frac{h(t_1)}{t_1} + \int_{t_1}^{t_2}\frac{h(s)}{s^2}\,\mathrm{d}s\right] \qquad (1 \le t_1 < t_2),$$

hence by (12.6.17),

$$g_B(t_2) - g_B(t_1) = \mathcal{O}\big(t_1^{-1-\varepsilon/2}\big) \qquad (1 \le t_1 < t_2).$$

Without loss of generality we assume: $\lim_{t\to\infty} g_B(t) = 0$, hence $g_B(t) = \mathcal{O}\big(t^{-1-\varepsilon/2}\big)$. Then g_B leads to an integrable contribution of order $\mathcal{O}(t^{-1-\frac{\varepsilon}{2}})$ in (12.6.21).

– The piecewise constant function g_A in (12.6.22) is increasing, because by Definition (12.6.8) of the Graf partition and (12.6.11),

$$\frac{q(t_0)}{t_0^{1-\varepsilon/2}} \in \Xi^{(\delta)}_{\mathcal{A}^-} \cap \Xi^{(\delta)}_{\mathcal{A}^+} \qquad \text{with } \mathcal{A}^\pm := \lim_{\gamma\searrow 0}\mathcal{A}(t_0 \pm \gamma) \text{ and } \mathcal{A}^+ \neq \mathcal{A}^-$$

satisfies the inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[J^E_{\mathcal{A}^+}\left(\frac{q(t)}{t^{1-\varepsilon/2}}\right) - J^E_{\mathcal{A}^-}\left(\frac{q(t)}{t^{1-\varepsilon/2}}\right)\right]\bigg|_{t=t_0} \ge 0\,.$$

So altogether, we can treat the case $g \neq 0$ in complete analogy to the case $g = 0$. We conclude that $\tilde{j}^E$ (and thus also $j^E = \tilde{j}^E + h$ and $j = j^I + j^E$) converges.

• Now if the limit $j^+ = 0$, then the asymptotic velocity exists: $\overline{v}^+(x_0) = 0$. So we now assume that $j^+ > 0$.

$Q(t) := q(t)/t$ satisfies $Q \in C^2([1,\infty), M)$. We now consider the set of cluster points of this curve:

$$\Omega := \left\{v \in M \;\middle|\; \exists (t_n)_{n\in\mathbb{N}} \text{ with } \lim_{n\to\infty} t_n = +\infty \text{ and } \lim_{n\to\infty} Q(t_n) = v\right\}.$$

The curve is contained in the sphere

$$S_{\mathcal{M}} := \{v \in M \mid \|v\|^2_{\mathcal{M}} = j^+\}\,.$$

As $S_{\mathcal{M}}$ is compact, Ω is not empty. We want to show that Ω consists of only one point (namely the asymptotic velocity $\overline{v}^+(x_0)$).

Recall the partition (12.6.4) of M and consider first a cluster point $v \in \Omega \cap \Xi_{\mathcal{C}}^{(0)}$ for which the number $|\mathcal{C}|$ of clusters is maximal. Define $U_\gamma(v) := \{v' \in M \mid \|v' - v\|_{\mathcal{M}} < \gamma\}$. As the atom $\Xi_{\mathcal{C}}^{(0)}$ of the partition is relatively open in the collision subspace $\Delta_{\mathcal{C}}^{(0)}$, for small $\gamma > 0$, the 2γ-neighborhood $U_{2\gamma}(v) \cap \Omega$ of v will lie in $\Xi_{\mathcal{C}}^{(0)}$.

For all $n \geq N(\gamma)$, one has $Q(t_n) \in U_{\gamma/2}(v)$, because v is a cluster point of $\big(Q(t_n)\big)_n$. We want to show that $Q(t) \in U_\gamma(v)$ for all $t \geq T(\gamma)$. As γ can be chosen arbitrarily small, this implies $\Omega = \{v\}$.

• Let $\chi \in C_c^\infty(M, [0,1])$ be a function with support in the neighborhood $U_{2\gamma}(v)$ whose restriction to $U_\gamma(v)$ is 1.

$$\Psi : [1,\infty) \to [0,\infty) \quad , \quad \Psi(t) = \chi\big(Q(t)\big) \left\| v_{\mathcal{C}}^E(t) - \frac{q_{\mathcal{C}}^E(t)}{t} \right\|_{\mathcal{M}}$$

is bounded and has the derivative

$$\begin{aligned}\Psi'(t) &= \chi\big(Q(t)\big) \Big\langle \Big(\tfrac{q_{\mathcal{C}}^E(t)}{t} - v_{\mathcal{C}}^E(t)\Big) \Big/ \Big\| \tfrac{q_{\mathcal{C}}^E(t)}{t} - v_{\mathcal{C}}^E(t) \Big\|_{\mathcal{M}}, \nabla V^E(q(t)) \Big\rangle \\ &\quad + \tfrac{1}{t} \Big\langle v(t) - \tfrac{q(t)}{t}, \nabla \chi\big(Q(t)\big) \Big\rangle \Big\| \tfrac{q_{\mathcal{C}}^E(t)}{t} - v_{\mathcal{C}}^E(t) \Big\|_{\mathcal{M}} - \tfrac{1}{t}\Psi(t)\,. \qquad (12.6.23)\end{aligned}$$

– Because of (12.6.20), the first term in the sum is in $L^1\big([1,\infty)\big)$.
– There exist such cutoff functions χ that have a product form

$$\chi(v) = \chi^I\big(\Pi_{\mathcal{C}}^I(v)\big)\ \chi^E\big(\Pi_{\mathcal{C}}^E(v)\big) \qquad (v \in M).$$

For large times t and $Q(t) \in U_{2\gamma}(v)$, it follows that $Q(t) \in \Delta_{\mathcal{C}}^{(\gamma/2)}$ (because otherwise v would not be a cluster point for the maximal number $|\mathcal{C}|$ of clusters).

Therefore $\chi^I\big(\Pi_{\mathcal{C}}^I(Q(t))\big) = 1$, and hence in the second term of (12.6.23),

$$\nabla\chi\big(Q(t)\big) = \nabla\chi^E\left(\Pi_{\mathcal{C}}^E(Q(t))\right) \in \Delta_{\mathcal{C}}^{(0)}\,.$$

So the absolute value of the second term is less than

$$\frac{C}{t} \left\| \frac{q_{\mathcal{C}}^E(t)}{t} - v_{\mathcal{C}}^E(t) \right\|_{\mathcal{M}}^2 \quad \text{, with} \quad C := \sup_x \|\nabla\chi(x)\|_{\mathcal{M}}\,.$$

From the integrability of the term $t \mapsto t^{-2}\int_1^t s^2 I_2(s)\,\mathrm{d}s$ in (12.6.21), one concludes by means of an integration by parts that $t \mapsto \frac{C}{t}\left\|\frac{q_{\mathcal{C}}^E(t)}{t} - v_{\mathcal{C}}^E(t)\right\|_{\mathcal{M}}^2$ also lies in $L^1\big([1,\infty)\big)$, and therefore so does the second term of (12.6.23).
– The third term of (12.6.23) can be written in the form

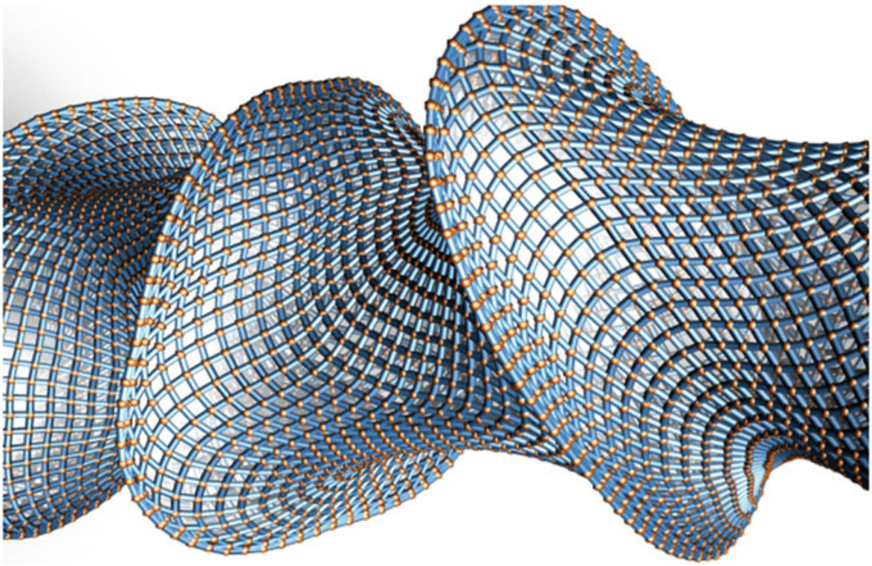
$$\frac{1}{t}\Psi(t) = \chi\big(Q(t)\big)\left\|\frac{\mathrm{d}}{\mathrm{d}t}\frac{q_C^E(t)}{t}\right\|_{\mathcal{M}}.$$

As it is non-negative, and the other two terms on the right hand side of (12.6.23) are integrable, and the left hand side of (12.6.23) is bounded, it follows that $t \mapsto \frac{1}{t}\Psi(t) \in L^1\big([1,\infty)\big)$.

• As we have therefore $\int_1^\infty \chi\big(Q(t)\big)\left\|\frac{\mathrm{d}}{\mathrm{d}t}\frac{q_C^E(t)}{t}\right\|_{\mathcal{M}} \mathrm{d}t < \infty$, and $Q(t_n) \in U_{\gamma/2}(v)$ for all sufficiently large n, it follows that $Q(t)$ remains in $U_\gamma(v)$ for t large. □

As mentioned on page 285, the asymptotic completeness, which we have just shown for bounded potentials, does not hold for all initial conditions for the Kepler potentials from celestial mechanics.

Chapter 13
Integrable Systems and Symmetries



K-surface (discrete surface of constant negative Gaussian curvature).
Image: courtesy of Ulrich Pinkall

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_13

13.1 What is Integrability? An Example

"When, however, one attempts to formulate a precise definition of integrability, many possibilities appear, each with a certain intrinsic theoretic interest."
D. BIRKHOFF, in: Dynamical Systems [Bi3]

In a heuristic meaning, a differential equation is integrable if we are able to 'write down' its solution.

There are two reasons why this "definition" obviously leaves much to be desired. For one thing, we would like to have a notion of integrability that tells something about the differential equation, rather than our mathematical abilities. On the other hand, it is not quite clear what is meant by 'write down'. Is the solution to be given in terms of 'known functions', in terms of convergent series, or possibly in terms of a limiting process?

We will proceed by first discussing an example of a Hamiltonian system that we consider as integrable, and thereafter we will abstract a notion of integrability from it.

13.1 Example (**Planar Motion in a Centrally Symmetric Potential**)
We consider the motion in a configuration space that is a plane, so the phase space will be $P := T^*\mathbb{R}^2_q \cong \mathbb{R}^4$ with its symplectic form $\omega_0 = dq_1 \wedge dp_1 + dq_2 \wedge dp_2$, and the Hamiltonian $H(p, q) := \frac{1}{2}\|p\|^2 + V(q)$, with a centrally symmetric potential

$$V \in C^\infty(\mathbb{R}^2_q, \mathbb{R}) \quad , \quad V(q) = W(\|q\|) .$$

The angular momentum $L : P \to \mathbb{R}$, $L(p, q) := -q_1 p_2 + q_2 p_1$ Poisson-commutes with H, i.e., $\{H, L\} = 0$; we abbreviate the *energy-angular momentum mapping* as

$$F := (H, L) \in C^\infty(P, \mathbb{R}^2) .$$

So we know that the orbit through $x_0 \in P$ stays in the set $F^{-1}(f) \subset P$ with $f := F(x_0)$. For regular values f in the image $F(P) \subset \mathbb{R}^2$ of phase space, this set is a 2-dimensional manifold.

A value $f = (h, \ell)$ is regular if and only if for all $x = (p, q) \in F^{-1}(f)$, the covectors $dH(x)$ and $dL(x)$ are linearly independent.

- In contrast, if $dH(x) = 0$, then one concludes from

$$dH(x) = p_1\, dp_1 + p_2\, dp_2 + \partial_{q_1} V(q)\, dq_1 + \partial_{q_2} V(q)\, dq_2 \tag{13.1.1}$$

  that $W'(\|q\|) = 0$ and $p = 0$, hence also $L(x) = 0$ and $H(x) = W(\|q\|)$. Thus the corresponding singular values are of the form $(h, \ell) = \big(W(\|q\|), 0\big)$, so they lie on the vertical axis (in Figure 13.1.1) of the (h, ℓ) plane.

- If however $dH(x) \neq 0$, but

$$dL(x) = c\, dH(x) \tag{13.1.2}$$

for some $c \in \mathbb{R}$, then one concludes from

$$dL(x) = q_2\, dp_1 - q_1\, dp_2 - p_2\, dq_1 + p_1\, dq_2$$

by comparing coefficients from (13.1.1) and (13.1.2) and letting $\mathbb{J} = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$:

$$q = c\, \mathbb{J} p \quad \text{and} \quad p = -c\, \mathbb{J} \nabla V(q)\, .$$

The set of such singular points is a null set in phase space P, and by $L(x) = \langle q, \mathbb{J} p\rangle = c\|p\|^2$, it corresponds to circular orbits. In Figure 13.1.1, the corresponding singular values can be recognized as the boundary curves of the set $F(P) \subset \mathbb{R}^2$.

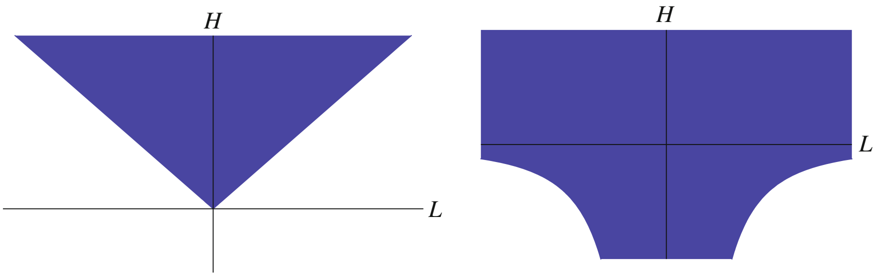


Figure 13.1.1 Energy-angular momentum mapping for centrally symmetric potentials. Left: Harmonic oscillator, right: Kepler potential

Due to the rotational symmetry of the Hamiltonian, polar coordinates (r, φ) are appropriate, where

$$q_1 = r \sin(\varphi) \quad , \quad q_2 = r \cos(\varphi)\, , \tag{13.1.3}$$

and we can enhance these to canonical coordinates. To this end, we make a generating function ansatz as in case 2 of Chapter 10.5 (page 238):

$$(p_r, p_\varphi, q_1, q_2) = \left(\frac{\partial S}{\partial r}, \frac{\partial S}{\partial \varphi}, \frac{\partial S}{\partial p_1}, \frac{\partial S}{\partial p_2}\right).$$

For $S(r, \varphi, p_1, p_2) := p_1 r \sin(\varphi) + p_2 r \cos(\varphi)$, Eq. (13.1.3) holds. Furthermore, we get then

$$p_r = p_1 \sin(\varphi) + p_2 \cos(\varphi) = \frac{\langle p, q\rangle}{\|q\|} \quad , \quad p_\varphi = p_1 r \cos(\varphi) - p_2 r \sin(\varphi) = L\, .$$

In terms of physics, we may simply view p_r as the radial component of the momentum $p = (p_1, p_2)$, whereas p_φ is the angular momentum.

Altogether, we obtain a canonical transformation

$$\Psi : T^*\left(\mathbb{R}^+ \times S^1\right) \to T^*\left(\mathbb{R}^2 \setminus \{0\}\right) \quad , \quad (p_r, p_\varphi, r, \varphi) \mapsto (p_1, p_2, q_1, q_2)\,.$$

Because of the identity $p_1^2 + p_2^2 = p_r^2 + \frac{L^2}{r^2}$ (obtained from trigonometric formulas), one has

$$H \circ \Psi\,(p_r, p_\varphi, r, \varphi) = \frac{p_r^2}{2} + \frac{p_\varphi^2}{2r^2} + W(r)\,.$$

As $H \circ \Psi$ does not depend on the φ coordinate, one gets $\dot{p}_\varphi = \dot{L} = 0$, hence $\ell := p_\varphi(0) = p_\varphi(t)$ is a constant of motion for $H \circ \Psi$.

$$K_\ell : T^*(\mathbb{R}^+) \to \mathbb{R} \quad , \quad K_\ell(p_r, r) = \tfrac{1}{2}p_r^2 + W_\ell(r) \quad \text{with} \quad W_\ell(r) := W(r) + \frac{\ell^2}{2r^2}$$

is a Hamiltonian with a single degree of freedom. W_ℓ is called the *effective potential*.

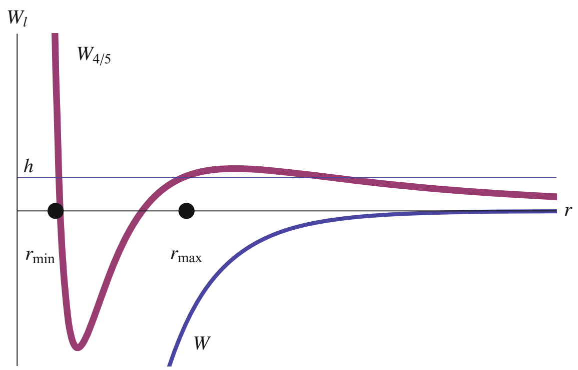
Since the energy $h := K_\ell\big(p_r(0), r(0)\big)$ is constant in time, we already know the orbits of the flow generated by K_ℓ.

Their time parametrization is obtained from $\dot{r} = p_r = \sqrt{2(h - W_\ell(r))}$ by inverting the integral

$$\pm \int_{r_0}^{r(t)} \frac{\mathrm{d}r}{\sqrt{2(h - W_\ell(r))}} = t - t_0\,.$$

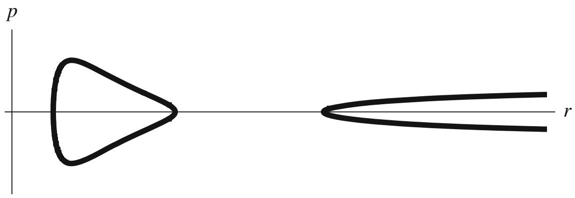
For $\ell \neq 0$, our assumption that V be continuous implies that $\lim_{r \searrow 0} W_\ell(r) = +\infty$, and therefore, for this energy h, the centrifugal potential $\frac{\ell^2}{2r^2}$ prevents the particle from visiting the origin. If instead we take a potential like $V(q) = -\frac{c}{\|q\|^\alpha}$ with $\alpha \in (0, 2)$, thus $W(r) = -\frac{c}{r^\alpha}$, the same conclusion still applies.

Under the hypothesis $V(q) \geq -c(1 + \|q\|^2)$, see Theorem 11.1, the existence of the solution to the Hamiltonian differential equation is guaranteed for all times, and the radial component of the orbit can be described by routine single variable calculus. Namely, only those values r for which $W_\ell(r) \leq h$ can occur; this condition is satisfied on certain intervals in $\mathbb{R}$.

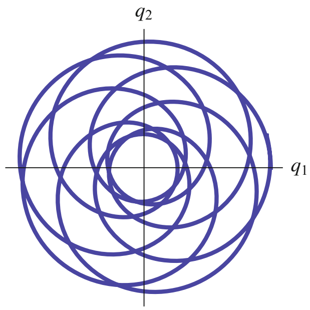


We only consider the interval that contains the initial value.

This interval is either of the form $[r_{\min}, r_{\max}]$ or else of the form $[r_{\min}, \infty)$, depending on whether we have bound or unbound motion. In the example of the *Yukawa* potential $W(r) := -\exp(-r)/r$, the energy shell in phase space looks as follows:



So for a regular value h, in the bound case, the radius corresponding to the location of the particle varies periodically in time between $r_{\min}$ and $r_{\max}$, whereas in the unbound case, it goes to infinity as $t \to \pm\infty$.



Plugging the solution $t \mapsto r(t)$ into the differential equation

$$\dot{\varphi} = \frac{\partial \tilde{H}}{\partial p_\varphi} = \frac{p_\varphi}{r^2} = \frac{\ell}{r^2(t)} \, ,$$

one obtains

$$\varphi(t) = \varphi(t_0) + \ell \textstyle\int_{t_0}^{t} r^{-2}(s)\, \mathrm{d}s \, .$$

In particular, the angular velocity is smallest near $r_{\max}$.

In the configuration plane (for the Yukawa potential), one obtains the form of motion described in the picture. In phase space, the connected components

$$M_{h,\ell} \subseteq F^{-1}(f) = \big\{(p,q) \in \mathbb{R}^4 \mid H(p,q) = h \, , \; L(p,q) = \ell\big\}$$

for regular values $f = (h, \ell)$ and bound motion correspond to a torus $\mathbb{T}^2 := S^1 \times S^1$, which is parametrized by φ and the time that has elapsed since the latest pericenter. In the case of unbound motion, these sets are cylinders $S^1 \times \mathbb{R}$.

We will see in Theorem 13.3 that these shapes of submanifolds are typical for integrable Hamiltonian systems. ◊

13.2 The Liouville-Arnol'd Theorem

We abstract the essential ingredients of the preceding example in a definition:

13.2 Definition *Let $H \in C^\infty(P, \mathbb{R})$ be a (Hamilton) function on the symplectic manifold (P, ω) of dimension $2n$.*

- *Then $F \in C^\infty(P, \mathbb{R})$ is called a* **constant of motion** *if $\{F, H\} = 0$.*
- *A set $\{F_1, \ldots, F_k\}$ of functions $F_i \in C^\infty(P, \mathbb{R})$ is said to be* **in involution** *if the Poisson brackets vanish:*

$$\{F_i, F_j\} = 0 \qquad \big(i, j \in \{1, \ldots, k\}\big).$$

- *The set is called* **independent** *if the set*

$$\{x \in P \mid dF_1(x) \wedge \ldots \wedge dF_k(x) = 0\}$$

  has Liouville measure[1] *zero.*
- *$\{F_1, \ldots, F_k\}$ is called (***Liouville-) integrable*** if the F_i are in involution and independent, and $k = n$.*
- *The function H is called* **integrable** *if there are, in addition to $F_1 := H$, another $n - 1$ constants of motion $F_2, \ldots, F_n$, such that $\{F_1, \ldots, F_n\}$ is integrable.*

At first, this definition of integrability appears to be somewhat abstract; however it permits to linearize the motion by introducing appropriate (semilocal) canonical coordinates

$$(I_1, \ldots, I_n\,,\ \varphi_1, \ldots, \varphi_n)\ .$$

In these coordinates, the Hamiltonian equations will take on the form

$$\dot{I}_k = 0 \quad , \quad \dot{\varphi}_k = \omega_k(I) \qquad \big(k \in \{1, \ldots, n\}\big)$$

so that their solution is simply

$$I_k(t) = I_k(0) \quad , \quad \varphi_k(t) = \varphi_k(0) + \omega_k(I)\, t\ .$$

We begin with a statement that has been developed into the following form over a period of more than 100 years (including the work by Henri Mineur [Min] from 1936).

---

[1] **Definition**: If (P, ω) is a symplectic manifold and $n := \frac{1}{2}\dim(P)$, then the measure on P that is induced by the volume form $\frac{(-1)^{\lfloor n/2 \rfloor}}{n!}\omega^{\wedge n}$ is called the *Liouville measure* (giving rise to an invariant measure on an energy surface, defined in Example 9.4.2. See also Remark 10.14). For $(P, \omega) = (T^*\mathbb{R}^n, \omega_0)$, it coincides with the Lebesgue measure λ^{2n} on $\mathbb{R}^{2n} \cong T^*\mathbb{R}^n$.

13.3 Theorem (**Liouville and Arnol'd**)
Let (P, ω) be a $2n$-dimensional symplectic manifold, and let $\{F_1, \dots, F_n\}$ be integrable. If $f \in F(P) \subset \mathbb{R}^n$ is a regular value[2] *of the mapping*

$$F := \begin{pmatrix} F_1 \\ \vdots \\ F_n \end{pmatrix} : P \to \mathbb{R}^n ,$$

then each compact connected component M_f of $F^{-1}(f)$ is diffeomorphic to an n-dimensional torus $\mathbb{T}^n = (S^1)^n$.
More specifically, there exist angle coordinates $\varphi_1, \dots, \varphi_n$ on M_f and frequencies $\omega_1, \dots, \omega_n \in \mathbb{R}$ in which the flow generated by H on M_f has the form

$$\varphi_k(t) = \varphi_k(0) + \omega_k\, t \pmod{2\pi} \quad , \quad \big(t \in \mathbb{R},\ k \in \{1, \dots, n\}\big). \tag{13.2.1}$$

Proof:

- We already know (Example 10.42.2) that M_f is an n-dimensional Lagrangian submanifold.
- Let $B_\varepsilon(f) := \{f' \in \mathbb{R}^n \mid \|f' - f\| \le \varepsilon\}$, the closed ball with small radius $\varepsilon > 0$; then the connected component K of M_f in the set $F^{-1}\big(B_\varepsilon(f)\big) \subset P$ is also compact, because M_f is compact and f is a regular value.
- Moreover, since $dF_k(X_{F_l}) = \{F_k, F_l\} = 0$, the vector fields X_{F_l} are tangential to the level sets of F by Theorem 10.16. By Theorem 3.27, the flow $\Phi_t^k : K \to K$ on K that is generated by the Hamiltonian vector field X_{F_k} for the k^{th} constant of motion F_k exists for all times $t \in \mathbb{R}$.
- We know, for all times $t_i \in \mathbb{R}$, that

  $$\Phi_{t_k}^k \circ \Phi_{t_l}^l = \Phi_{t_l}^l \circ \Phi_{t_k}^k \qquad \big(k, l \in \{1, \dots, n\}\big),$$

  because by Theorem 10.21, the flows generated by vector fields X, Y commute if and only if $[X, Y] = 0$; and by Remark 10.23, the commutator $[X_{F_k}, X_{F_l}] = -X_{\{F_k, F_l\}} = 0$, since $\{F_k, F_l\} = 0$. Now we want to use the mapping

  $$\Psi : \mathbb{R}^n \times M_f \to M_f \quad , \quad \big((t_1, \dots, t_n), x\big) \mapsto \Phi_{t_1}^1 \circ \dots \circ \Phi_{t_n}^n(x) \tag{13.2.2}$$

  for a parametrization of M_f. This is indeed a mapping into M_f, because the flows Φ_t^k leave M_f invariant (as their vector fields X_{F_k} are tangential to M_f).
  Moreover, $\Psi_t : M_f \to M_f \quad , \quad \Psi_t(x) := \Psi(t, x)$ satisfies

  $$\Psi_0 = \mathrm{Id}_{M_f} \quad , \quad \Psi_s \circ \Psi_t = \Psi_{s+t} \qquad \big(s, t \in \mathbb{R}^n\big).$$

[2] Equivalently: the functions $F_1, \dots, F_n$ are independent on the level set $F^{-1}(f)$, i.e., $dF_1 \wedge \dots \wedge dF_n(x) \neq 0$ if $F(x) = f$.

So Ψ is a group action of the Lie group $\mathbb{R}^n$ on M_f. (See Remark 2.17.7 and Definition E.4).

- As the dF_k are independent, so are the n vector fields X_{F_k} on M_f. Therefore (also using the compactness of M_f), there exists a neighborhood $U \subset \mathbb{R}^n$ of zero such that for all $x \in M_f$, the mapping $U \to M_f$, $t \mapsto \Psi(t, x)$ is a diffeomorphism onto its image $\Psi\big(U, \{x\}\big)$ (see Figure 13.2.1).

So the group action is *locally free*. The image $\Psi\big(\mathbb{R}^n, \{x\}\big)$ of the mapping $\mathbb{R}^n \to M_f$, $t \mapsto \Psi(t, x)$ is therefore open in M_f (and nonempty!). But its complement $M_f \setminus \Psi\big(\mathbb{R}^n, \{x\}\big)$ is also open. For if y is in the complement of the image, then the neighborhood $\Psi\big(U, \{y\}\big)$ of y must also be in the complement, because otherwise there would exist some $z \in \Psi\big(\mathbb{R}^n, \{x\}\big) \cap \Psi\big(U, \{y\}\big)$, hence $z = \Psi_t(x) = \Psi_s(y)$ with appropriate $s, t \in \mathbb{R}^n$, but this would imply $y = \Psi_{t-s}(x)$.

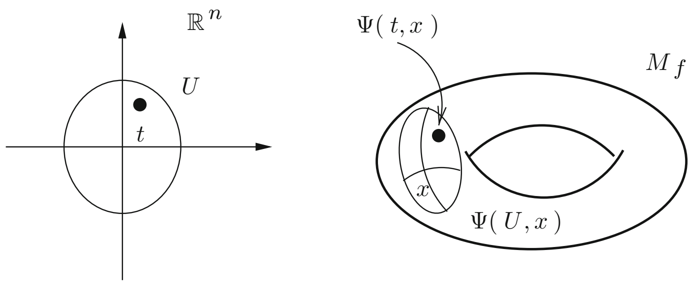


Figure 13.2.1 Group action Ψ of $\mathbb{R}^n$ on the compact connected component M_f of $F^{-1}(f)$

So, since $\Psi\big(\mathbb{R}^n, \{x\}\big)$ is open and closed and nonempty, it must be a connected component of M_f. But by assumption, M_f is connected, so one has $\Psi\big(\mathbb{R}^n, \{x\}\big) = M_f$; we say the group action is *transitive*.

- The *isotropy group* Γ of a point $x \in M_f$ (see figure below) is defined by

$$\Gamma \equiv \Gamma_x := \{t \in \mathbb{R}^n \mid \Psi_t(x) = x\}\,. \tag{13.2.3}$$

Γ is independent of the choice of x, because by the transitivity of Ψ, for any $y \in M_f$, there exists $s \in \mathbb{R}^n$ with $y = \Psi_s(x)$. As Ψ is an action of the abelian group $\mathbb{R}^n$, we infer that $\Gamma_y = \{t \in \mathbb{R}^n \mid \Psi_{t+s}(x) = \Psi_s(x)\} = \Gamma_x$.

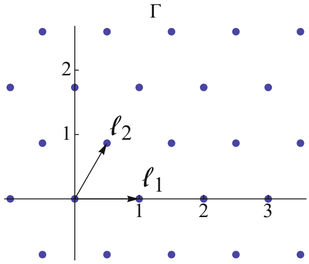
As the group action is locally free, there exists a neighborhood $U \subset \mathbb{R}^n$ of $0 \in \Gamma$ for which $U \cap \Gamma = \{0\}$. So if $t \in \Gamma$, then t is the only point in the neighborhood $U + t$ that belongs to the isotropy group. Therefore Γ is a *discrete* subgroup of $\mathbb{R}^n$ (viewed as a subset of the topological space $\mathbb{R}^n$, see Def. A.20). We temporarily suspend the proof of Theorem 13.3 to study these subgroups.

13.4 Lemma *Let $\Gamma \subset \mathbb{R}^n$ be a discrete subgroup and $k := \dim\big(\mathrm{span}_{\mathbb{R}}(\Gamma)\big)$. Then there exist linearly independent vectors $\ell_1, \ldots, \ell_k \in \Gamma$ such that*

$$\Gamma = \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_k) := \left\{ \textstyle\sum_{i=1}^{k} z_i \ell_i \,\middle|\, z_i \in \mathbb{Z} \right\}.$$

Proof: We call a subspace $U \subset \mathbb{R}^n$ a Γ-*subspace*, if

$$\mathrm{span}_{\mathbb{R}}(U \cap \Gamma) = U\,,$$



and we construct inductively bases $\ell_1, \ldots, \ell_m \in \Gamma$ for appropriate Γ-subspaces $U_m \subset \mathbb{R}^n$ of respective dimensions $m \le k$, bases that will at the same time be $\mathbb{Z}$-*bases* of $U_m \cap \Gamma$, i.e.,

$$U_m \cap \Gamma = \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_m).$$

- $\{0\} \subset \Gamma$ is a 0-dimensional Γ-subspace.
- If for $m < k$, we have constructed a Γ-subspace U_m with basis $\ell_1, \ldots, \ell_m$, then there exists a vector $\ell_{m+1} \in \Gamma$ that is linearly independent of that basis and whose distance $a > 0$ from the subspace $\mathrm{span}_{\mathbb{R}}(\ell_1, \ldots, \ell_m)$ is minimal.[3]
  This is because $\Gamma \setminus \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_m) \neq \emptyset$, and from any ℓ_{m+1} in this set, we can obtain another vector $\ell_{m+1} - \ell$ by translation with $\ell \in \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_m)$; this new vector has the same distance a, but its length is bounded as $\|\ell_{m+1} - \ell\| \le a + \sum_{i=1}^{m} \|\ell_i\|$. If the infimum of the distances a to the subspace were zero, we would get a contradiction to the discreteness of Γ. A compactness argument guarantees that the infimum is a minimum.
- By construction, $U_{m+1} = \mathrm{span}_{\mathbb{R}}(\ell_1, \ldots, \ell_{m+1})$ is a Γ-subspace.
- $\ell_1, \ldots, \ell_{m+1}$ is a $\mathbb{Z}$-basis of U_{m+1}, because every $\ell \in U_{m+1} \cap \Gamma$ can be uniquely written in the form

$$\ell = z\ell_{m+1} + u \quad \text{with } u \in U_m \text{ and } z \in \mathbb{R}\,.$$

  The vector $\ell - \lfloor z \rfloor \ell_{m+1} \in U_{m+1} \cap \Gamma$ has distance $(z - \lfloor z \rfloor) \cdot a$ from the subspace U_m. As this distance cannot lie strictly between 0 and a, we must have $z \in \mathbb{Z}$. Therefore $u \in \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_m)$, and ultimately $\ell \in \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_{m+1})$. □

---

[3]If one were to take instead a *shortest* vector $\ell_{m+1} \in \Gamma$ that is linearly independent of $\ell_1, \ldots, \ell_m$, this would in general not work in high dimensions, see for example QUAISSER [Qu], §5.1.

Conclusion of the Proof of Theorem 13.3:

- If we extend the $\mathbb{Z}$-basis $\ell_1, \dots, \ell_k$ from Lemma 13.4 to an $\mathbb{R}$-basis $\ell_1, \dots, \ell_n$ of $\mathbb{R}^n$, then a change of basis yields the group isomorphism and diffeomorphism

$$\mathbb{R}^n/\Gamma \cong \mathbb{R}^n/\mathbb{Z}^k \cong \mathbb{R}^{n-k} \times (\mathbb{R}/\mathbb{Z})^k \cong \mathbb{R}^{n-k} \times \mathbb{T}^k .$$

  As M_f is compact by hypothesis, and $M_f \cong \mathbb{R}^n/\Gamma$, it follows that $k = n$ and $M_f \cong \mathbb{T}^n$.
- With M_f thus being diffeomorphic to the n-torus, it follows that points on M_f can be parametrized by n angles $\varphi_1, \dots, \varphi_n$. This can be done even in such a way that the flow generated by $H = F_1$ attains a particularly simple form: Let $\ell_1, \dots, \ell_n$ again be a basis of the isotropy group $\Gamma \subset \mathbb{R}^n$ from (13.2.3). Then for every point $x \in M_f$, the mapping

$$[0, 2\pi[^n \longrightarrow M_f \quad , \quad (\varphi_1, \dots, \varphi_n) \longmapsto \Psi\left(\sum_{i=1}^{n} \frac{\varphi_i \ell_i}{2\pi}, x\right), \tag{13.2.4}$$

  which is defined by restriction of the $\mathbb{R}^n$-action (13.2.2), is bijective. Assume the vector $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$ is represented as $e_1 = \sum_{i=1}^{n} \frac{\omega_i}{2\pi} \ell_i$. Then the flow generated by H is of the form

$$\Phi^1_{t_1}(y) = \Psi\big((t_1, 0, \dots, 0),\ y\big) = \Psi\left(\sum_{i=1}^{n} t_1 \frac{\omega_i \ell_i}{2\pi},\ y\right).$$

  Therefore, the coordinates $\varphi_1(t), \dots, \varphi_n(t)$ of $\Phi^1_t(y)$ satisfy Eq. (13.2.1). □

13.5 Remark
A motion on an n-torus that is as given in the second part of the theorem is called a *conditionally periodic motion*, see also page 221 and page 395 (Figure 13.2.2). ◊

13.6 Remark (**Symplectic Integrators**)
Most Hamiltonian systems are not integrable. One therefore has to resort to numerical methods to calculate their dynamics.

Ironically, integrable systems are useful in developing numerical methods that are adapted to the problem.

Namely, it is advantageous for a numerical approximation to Hamiltonian dynamics, if this approximation itself produces a symplectic time evolution. For instance, Theorem 7.3 about strong stability tells us that in the linear case, small *symplectic* perturbations of a nonresonant Hamiltonian system will not violate its Lyapunov-stability, whereas the same is not true for general perturbations.

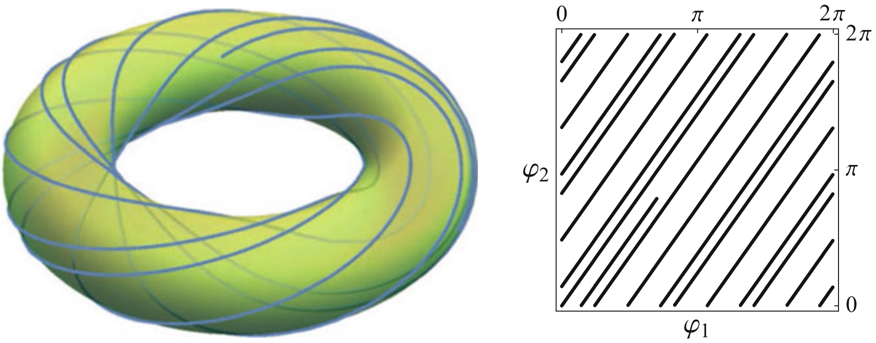


Figure 13.2.2 Conditionally periodic motion with frequency ratio $\sqrt{2}$ on the torus $\mathbb{T}^2$

The corresponding numerical methods are called symplectic integrators.

The basic idea can be illustrated in the example of motion in a potential, with the Hamiltonian $H(p,q) = T(p) + V(q)$ on the phase space $P := \mathbb{R}^d_p \times \mathbb{R}^d_q$. The terms T and V in the sum, viewed as Hamiltonians themselves, are integrable, and they have complete flows

$$\Phi_t(p,q) = \big(p, q + t\nabla T(p)\big) \text{ and } \Psi_t(p,q) = \big(p - t\nabla V(q), q\big) \quad \big(t \in \mathbb{R},\ (p,q) \in P\big).$$

Let us assume that the flow generated by H is complete as well, so that we can denote it as $\Xi_t : P \to P$ with $t \in \mathbb{R}$. An arbitrary composition $\Phi_{c_1 t} \circ \Psi_{d_1 t} \circ \ldots \circ \Phi_{c_k t} \circ \Psi_{d_k t}$ of the flows is symplectic. We call such a composition a symplectic integrator of n^{th} *order*, if

$$\Phi_{c_1 t} \circ \Psi_{d_1 t} \circ \cdots \circ \Phi_{c_k t} \circ \Psi_{d_k t} = \Xi_t + \mathcal{O}(t^{n+1}),$$

possibly with reference to a metric on the phase space that is adjusted to the problem. For example, the combination $c_1 = d_1 = 1$ yields a symplectic integrator of the first order; and

$$\Phi_{t/2} \circ \Psi_t \circ \Phi_{t/2} = \Xi_t + \mathcal{O}(t^3),$$

etc. see YOSHIDA [Yo]. Numerically, the concatenated mappings are iterated $\mathcal{O}(T/t)$ times to obtain an error of order $\mathcal{O}(t^n)$ after a total time T. ♢

13.7 Literature As indicated by the quotation from Birkhoff on page 326, there are many notions of integrability that can be related to each other, but refer to different areas of applications.

In particular, nonlinear *partial* differential equations can be integrable. An example is the Korteweg-de Vries equation $u_t = 6uu_x - u_{xxx}$ discussed by ABRAHAM and MARSDEN in [AM, Example 5.5.7]. Its Hamiltonian perturbation theory is discussed, e.g., in the book [KP] by KAPPELER and PÖSCHEL. OLSHANETSKY and PERELOMOV review relations to Lie theory in [OP].

An overview of integrability is presented in [BBT]. Algebraic aspects can be found for instance in ARNOL'D and NOVIKOV [AN], and in MOSER [Mos2].

Differential geometry allows for a deeper understanding of integrability. But also objects of differential geometry, like surfaces, can be solutions to integrable partial differential equations. A discretized version of such a surface is shown at the beginning of the chapter, on page 325, and discrete differential geometry of such structures is studied in [BoSu] by BOBENKO and SURIS. ◊

13.3 Action-Angle Coordinates

In a next step, we will introduce angle coordinates not just on a single torus M_f, but extend these to an open neighborhood of M_f in phase space, and complement them to canonical coordinates in the sense of Remark 10.18 by means of n coordinates that are called action coordinates.

13.8 Theorem (**Action-Angle Coordinates**)
Under the hypotheses of the Liouville-Arnol'd theorem (Theorem 13.3), every compact connected component M_f of $F^{-1}(f)$ has a neighborhood $M_f \subseteq U \subseteq P$, with action coordinates $I_k : U \to \mathbb{R}$ and angle coordinates $\varphi_k : U \to \mathbb{R}/(2\pi\mathbb{Z})$, for $k = 1, \ldots, n$, such that

- *the symplectic form is in canonical format $\omega = \sum_{k=1}^{n} d\varphi_k \wedge dI_k$,*
- *and the Hamiltonian differential equations are in the following form:*

$$\dot{I}_k = 0 \quad , \quad \dot{\varphi}_k = \omega_k(I) \qquad (k = 1, \ldots, n).$$

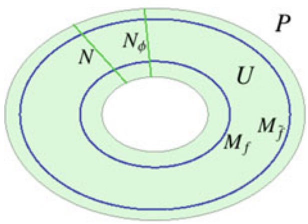
Proof:

- The fact that f is a regular value of F is in itself not sufficient to guarantee the existence of an open neighborhood $\tilde{U} \subseteq \mathbb{R}^n$ of f that also consists of regular values of F only.[4] If however U is chosen as that connected component of $F^{-1}(\tilde{U}) \subseteq P$ which contains M_f, then due to the compactness of M_f, one can guarantee by shrinking $\tilde{U}$ that all $x \in U$ are regular points of F.

- The common zeros of the n angle coordinates are to lie on a certain n-dimensional imbedded surface

$$N : \tilde{U} \to U \quad \text{with} \quad F \circ N = \mathrm{Id}_{\tilde{U}} \qquad (13.3.1)$$

  that is yet to be determined precisely, and that is to be transversal to the tori $M_{\tilde{f}}$ $(\tilde{f} \in \tilde{U})$.



[4]**Example**: The value $f = 0$ of $F : (0, \infty) \to \mathbb{R}$, $x \mapsto x \sin(1/x)$ is regular, but every neighborhood of f contains extremal values of F.

Since F is regular on U, the mapping

$$\Psi : \mathbb{R}^n \times U \to U \quad , \quad \big((t_1, \ldots, t_n), x\big) \mapsto \Phi^1_{t_1} \circ \cdots \circ \Phi^n_{t_n}(x) \,, \tag{13.3.2}$$

which extends the mapping (13.2.2) from M_f to U, is a locally free group action of $\mathbb{R}^n$. In analogy to (13.2.3), the isotropy group

$$\Gamma_x := \{t \in \mathbb{R}^n \mid \Psi_t(x) = x\} \tag{13.3.3}$$

of a point $x \in U$ depends only on $F(x)$. Since by the implicit function theorem, the points of the lattice Γ_x depend smoothly on x, and since we may assume that $\tilde{U}$ is simply connected, we can choose a $\mathbb{Z}$-basis $\ell_1(x), \ldots, \ell_n(x) \in \mathbb{R}^n$ of Γ_x whose vectors depend smoothly on $x \in U$. Then there also exists a section (13.3.1) in the $\mathbb{T}^n$-bundle $F : U \to \tilde{U}$.
Generalizing (13.2.4), every point $y \in U$ can be uniquely represented in the form

$$y = \Psi\left(\sum_{i=1}^{n} \frac{\varphi_i \ell_i(x)}{2\pi}, x\right) \text{ with } \varphi_i \in [0, 2\pi),\, x := N \circ F(y) \text{ and } \ell_i(x) = \ell_i(y)\,. \tag{13.3.4}$$

Thus we have defined angle coordinates $\varphi_1, \ldots, \varphi_n$ on U. While these are discontinuous, their exterior derivatives $d\varphi_k$ extend smoothly.

- If the symplectic form on U is to be in the canonical format $\omega = \sum_{k=1}^n dI_k \wedge d\varphi_k$, then it is in particular necessary that the Poisson brackets $\{\varphi_i, \varphi_k\}$ vanish (according to Remark 10.18). Because of the relation $\{\varphi_i, \varphi_k\} = L_{X_{\varphi_k}} \varphi_i$, this holds on N if and only if the locally Hamiltonian vector fields X_{φ_k} are tangential to the manifold N, i.e., in view of $\{\varphi_i, \varphi_k\} = \omega(X_{\varphi_i}, X_{\varphi_k})$, if N is a Lagrangian submanifold.
  Otherwise we rename the section N from (13.3.1) into $\hat{N}$. The deviation from the property of being Lagrangian is measured by the closed 2-form $\hat{N}^*\omega$ on $\tilde{U}$. We assume for example that $\tilde{U}$ is star-shaped (for instance a ball) and write $\hat{N}^*\omega$ as an exterior derivative $d\hat{\alpha}$ of a 1-form $\hat{\alpha}$ on $\tilde{U}$; this can be done because of the Poincaré-Lemma B.45. Lifting this 1-form to U, i.e., letting $\alpha := F^*(\hat{\alpha})$, we get a vector field $X : U \to TU$ that is tangental to the tori by $\mathbf{i}_X\omega = \alpha$. The Lie derivative of the symplectic form with respect to X is $L_X\omega = d\mathbf{i}_X\omega = d\alpha$. Consequently, the time-1 flow $\Phi : U \to U$ of $-X$ maps the submanifold $\hat{N}$ onto the Lagrangian submanifold $N := \Phi(\hat{N})$. This latter can also be viewed as a section (13.3.1), because $F \circ \Phi = F$.
  So whereas the submanifold $N = N_0$ corresponds to the value $\varphi = 0$, the level sets $N_{\tilde{\varphi}}$ of the angle variables are obtained from this submanifold by applying the Hamiltonian flow generated by the action variables $I = \tilde{I} \circ F$ with time $\tilde{\varphi}$. So far, we have not determined whether indeed we can find $\tilde{I}$ in such a way that $\{\varphi_i, I_k\} = \delta_{i,k}$ holds. In any case, for *any* choice of $\tilde{I}$, the $N_{\tilde{\varphi}}$ will be Lagrangian manifolds, so the Poisson brackets $\{\varphi_i, \varphi_k\}$ vanish on U.

- To find action coordinates $I_k : U \to \mathbb{R}$ associated with the angles φ_k, we are looking for a property of the mapping

$$I = (I_1, \ldots, I_n) : U \to \mathbb{R}^n$$

that determines for I to be appropriate. I only depends on the values of F, so $I = \tilde{I} \circ F$, with $\tilde{I} : \tilde{U} \to \mathbb{R}^n$. Since $\{F_i, F_k\} = 0$, the I_k will also Poisson-commute.
The derivative $\mathrm{D}\tilde{I}$ is determined by the intended values $\{\varphi_i, I_k\} = \delta_{i,k}$ of the Poisson brackets. The first argument of Ψ in (13.3.4) is equal to $t(\varphi) := \sum_{i=1}^{n} \frac{\varphi_i \ell_i}{2\pi}$. So using the dual basis vectors $\ell_i^* : U \to \mathbb{R}^n$, $\langle \ell_i^*, \ell_j \rangle := 2\pi\delta_{i,j}$, it follows that $\langle \ell_i^*, t(\varphi) \rangle = \varphi_i$. Finally, from

$$\{\varphi_i, F_j\} = L_{X_{F_j}} \varphi_i = L_{X_{F_j}} \langle \ell_i^*, t(\varphi) \rangle = \left\langle \ell_i^*, L_{X_{F_j}} t(\varphi) \right\rangle = \langle \ell_i^*, e_j \rangle = \ell_{i,j}^* \,, \tag{13.3.5}$$

it follows that

$$\delta_{i,k} = \{\varphi_i, I_k\} = \{\varphi_i, \tilde{I}_k \circ F\} = \sum_{j=1}^{n} \{\varphi_i, F_j\} \, \mathrm{D}_j \tilde{I}_k \circ F = \left\langle \ell_i^*, \mathrm{D}\tilde{I}_k \circ F \right\rangle ,$$

or $\mathrm{D}\tilde{I}_k \circ F = \frac{\ell_k}{2\pi}$. As the basis vectors ℓ_k are constant on the tori, and can therefore be written as $\tilde{\ell}_k \circ F$, the determining equation is simply

$$\mathrm{D}\tilde{I} = \frac{\tilde{L}}{2\pi} \,. \tag{13.3.6}$$

where $\tilde{L} := (\tilde{\ell}_1, \ldots, \tilde{\ell}_n)$.
- For $n > 1$ dimensions, the mapping $\tilde{L} : \tilde{U} \to \mathrm{Mat}(n, \mathbb{R})$ must satisfy an integrability condition in order for (13.3.6) to have a solution. Namely, by the commutability of partial derivatives of $\tilde{I}$, one must have

$$\mathrm{D}_i \tilde{\ell}_{j,k} = \mathrm{D}_k \tilde{\ell}_{j,i} \qquad (i, j, k = 1, \ldots, n). \tag{13.3.7}$$

Conversely, by the Poincaré lemma (Theorem B.48), Eq. (13.3.6) is solvable on simply connected domains $\tilde{U} \subseteq \mathbb{R}^n$, if condition (13.3.7) holds.
From the Poisson identity

$$\big\{\varphi_i, \{\varphi_j, F_k\}\big\} + \big\{F_k, \{\varphi_i, \varphi_j\}\big\} + \big\{\varphi_j, \{F_k, \varphi_i\}\big\} = 0 \,, \tag{13.3.8}$$

we conclude with $\{\varphi_i, \varphi_j\} = 0$ and (13.3.5) that $\{\varphi_i, \ell_{j,k}^*\} = \{\varphi_j, \ell_{i,k}^*\}$. Denoting the inverse matrix of $\tilde{L}$ by $\tilde{M} : \tilde{U} \to \mathrm{Mat}(n, \mathbb{R})$, its entries $\tilde{m}_{j,k}$ satisfy the relation $\tilde{\ell}_{i,j} \tilde{m}_{j,k} = \delta_{i,k}$ (with the Einstein summation convention in effect), and $\{\varphi_i, \tilde{m}_{k,j} \circ F\} = \{\varphi_j, \tilde{m}_{k,i} \circ F\}$.

Using (13.3.5) one more time, we obtain from (13.3.8) the relation

$$\tilde{m}_{r,i}\, \mathrm{D}_r \tilde{m}_{k,j} = \tilde{m}_{r,j}\, \mathrm{D}_r \tilde{m}_{k,i} \quad \text{or} \quad \mathrm{D}_i \tilde{m}_{a,b} = \tilde{\ell}_{s,i}\, \tilde{m}_{r,b}\, \mathrm{D}_r \tilde{m}_{a,s}\,. \tag{13.3.9}$$

Because $\mathrm{D}_i\big(\tilde{M}\tilde{L}\big) = 0$, one has $\mathrm{D}_i\, \tilde{\ell}_{j,k} = -\,\tilde{\ell}_{j,a}\, \tilde{\ell}_{b,k}\, \mathrm{D}_i \tilde{m}_{a,b}$. Plugging (13.3.9) into the right hand side confirms the integrability condition (13.3.7):

$$\mathrm{D}_i\, \tilde{\ell}_{j,k} = -\,\tilde{\ell}_{j,a}\, \tilde{\ell}_{b,k}\, \tilde{\ell}_{s,i}\, \tilde{m}_{r,b}\, \mathrm{D}_r \tilde{m}_{a,s} = -\,\tilde{\ell}_{j,a}\, \tilde{\ell}_{s,i}\, \mathrm{D}_k \tilde{m}_{a,s} = \mathrm{D}_k\, \tilde{\ell}_{j,i}\,. \qquad \Diamond$$

By the way: On the domain U, the vector valued function $I : U \to \mathbb{R}^n$ of action variables is determined by the derivative (13.3.6) up to addition of an arbitrary vector from $\mathbb{R}^n$.

13.9 Remark (**Relevance of the Action-Angle Coordinates**)
Why are we interested in the action variables I at all, rather than being content with the constants of motion F?
The reason is that they are so simple that the underlying geometry of the phase space emerges clearly:

1. On one hand, the Hamiltonian flow generated by I_k on M_f only changes the k^{th} angle φ_k. In contrast, the flow generated by F_k will in general change all angles simultaneously.
2. Even if it is true that $\{\varphi_i, F_k\} = 0$ for all $i \neq k$ (in particular in one degree of freedom, $n = 1$), the F_k, together with "appropriate" angle coordinates φ_k will in general not be canonical coordinates, i.e., the Poisson brackets $\{\varphi_k, F_i\}$ are not equal to $\delta_{i,k}$, in contrast to $\{\varphi_k, I_i\}$.

Due to their simplicity, the action-angle coordinates are an ideal starting point for the study of Hamiltonian perturbations (which are mostly not integrable), see Chapter 15. $\Diamond$

In the following section, we present a method by which the action coordinates can be found by means of integration.[5] So it is in this sense that the integrability of a Hamiltonian system indeed allows us to to calculate its solutions.

Now how do we construct the action variables from the constants of motion F_k? We will use a 1-form θ with its corresponding symplectic 2-form $\omega = -d\theta$. On cotangent spaces $(P, \omega) = (T^*N, \omega_0)$, the tautological 1-form θ_0 is a natural choice. But also on the phase space neighborhood $U \subset P$ from Theorem 13.8, there is always such a θ, namely by construction of the action-angle coordinates for example $\sum_{k=1}^n I_k \wedge d\varphi_k$.

Now for a given point $x_0 \in M_f$, we attempt to define a function on M_f by

$$S(x) := \int_{x_0}^{x} \theta\restriction_{M_f}\,.$$

[5]sometimes called by *quadratures*, since the area of a domain delimited by a curve is to be found.

This expression is to be understood as the line integral of the symplectic 1-form θ along a path $\gamma : [0, 1] \to M_f,\ \gamma(0) = x_0,\ \gamma(1) = x$ connecting x_0 to x. Does now S, as suggested by the notation, depend only on x, but not on the path γ?

To this end consider two paths γ_0 and $\gamma_1 : [0, 1] \to M_f$ with the given endpoints. First we will assume that γ_0 can be deformed into γ_1 while keeping the endpoints fixed, i.e., we assume the existence of a continuous *homotopy* $H : I \times I \to M_f$, $I := [0, 1]$ with

$$H(t, 0) = \gamma_0(t) \quad , \quad H(t, 1) = \gamma_1(t) \quad , \quad H(0, y) \equiv x_0 \quad , \quad H(1, y) \equiv x \, ,$$

see Appendix A.22. Then indeed, $\int_{\gamma_0} \theta\restriction_{M_f} = \int_{\gamma_1} \theta\restriction_{M_f}$, because

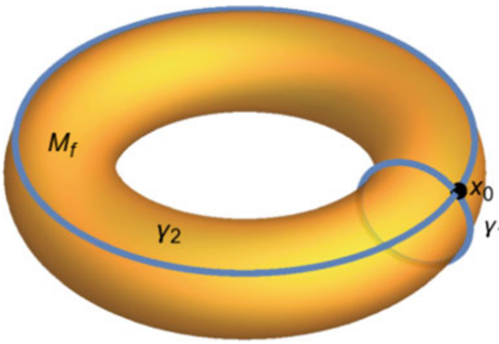
$$\int_{\gamma_0} \theta\restriction_{M_f} - \int_{\gamma_1} \theta\restriction_{M_f} = \int_{\partial(I\times I)} H^*(\theta) = \int_{I\times I} dH^*(\theta) = \int_{I\times I} H^*(d\theta) = 0$$

by the Stokes theorem (Theorem B.39), and because the symplectic 2-form $\omega = -d\theta$ vanishes identically on the Lagrangian submanifold M_f.

It looks as if we had proved that $S(x)$ does not depend on the path; but beware! We have assumed that the paths are homotopic. But not all paths on the torus are homotopic to each other. In particular, we can find n paths $\gamma_1, \ldots, \gamma_n$ with $\gamma_i(0) = \gamma_i(1) = x_0$, that *only* intersect at x_0. Namely γ_i traverses precisely the i^{th} angle once in positive direction. So we can calculate the integral '$S(x_0)$' in many different ways, say, by interpreting it as a notation for $\int_{\gamma_k} \theta$. Let us try and denote these integrals as

$$I_k := \frac{1}{2\pi} \int_{\gamma_k} \theta \qquad (k = 1, \ldots, n). \tag{13.3.10}$$

Are the I_k zero? In general, they are not.



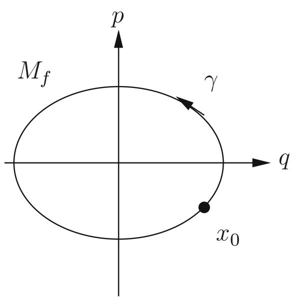


Figure 13.3.1 Left: Basis $\gamma_1, \ldots, \gamma_n$ of the fundamental group of the n-torus M_f. Right: Invariant 1-torus $M_f = \gamma([0, 1])$

13.10 Example In the particularly simple case of the symplectic phase space $P := \mathbb{R}_p \times \mathbb{R}_q$ with $\omega_0 = dq \wedge dp$, the level set M_f is the image of a closed curve $\gamma : [0, 1] \to P$ (see Figure 13.3.1, right). The integral from (13.3.10) is

$$I = \frac{1}{2\pi} \int_\gamma p\, dq = \frac{-1}{2\pi} \int_F \omega_0 \,,$$

where, by the Stokes theorem (Theorem B.39), $F \subset \mathbb{R}^2$ is the compact part of area that is enclosed by the circle $M_f = \partial F$. So I is proportional to the enclosed area. ♢

We can introduce paths $\gamma_1, \ldots, \gamma_n$ not only on M_f, but also on neighboring tori $M_{\tilde{f}}$ (with $\|\tilde{f} - f\|$ small), and by what we have just proved, the expressions I_k will only be functions $I_k = \tilde{I}_k(F_1, \ldots, F_n)$ of the constants of motion.[6]

The I_k are our candidates for action variables.

To introduce the angle variables, we consider the local representation of M_f as the graph of a function $p = p(I, q)$ with $I = \tilde{I}(f)$, see Figure 13.3.2, in the neighborhood of a typical point $x_0 = (p_0, q_0) \in M_f$. Of course this representation will not be possible for **all** $x_0 \in M_f$.

We let

$$\varphi_k(q) := \sum_{l=1}^{n} \int_{q_0}^{q} \frac{\partial p_l(I, q)}{\partial I_k} dq_l = \frac{\partial \tilde{S}}{\partial I_k} \tag{13.3.11}$$

where $\tilde{S}$ is the function (defined for a simply connected subset of the torus)

$$\tilde{S}(I, q) := \int_{q_0}^{q} p(I, q) \cdot dq \,.$$

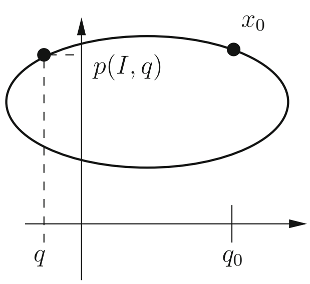


Figure 13.3.2 Local representation of M_f as the graph of a function

[6]By the way: To every loop $\gamma : S^1 \to M_f$ there corresponds a loop $\tilde{\gamma} : S^1 \to \Lambda(n)$, $\tilde{\gamma}(t) := T_{\gamma(t)} M_f$ in the Lagrange-Grassmann manifold. Thus as in Exercise 6.62, γ also gives rise to a Maslov index.

So $p = \frac{\partial \tilde{S}}{\partial q}$ and $\varphi = \frac{\partial \tilde{S}}{\partial I}$, hence the transformation $(p, q) \mapsto (I, \varphi)$ is locally canonical.

Now the local coordinates φ_k can indeed be interpreted as angles, i.e., we can interpret $\exp(i\varphi_k)$ as a function $M_f \to S^1 \subset \mathbb{C}$. This is because in a neighborhood of $x_0 \in M_f$, the expression

$$\exp\bigl(i\varphi_k(x)\bigr) = \exp\left(i \frac{\partial}{\partial I_k} \int_{x_0}^{x} \theta\restriction_{M_f}\right) \qquad (x \in M_f) \tag{13.3.12}$$

coincides with Definition (13.3.11), and by (13.3.10), the argument of the exponential function in (13.3.12), modulo $2\pi i$, is independent of the choice of path $\gamma : I \to M_f$ between $\gamma(0) = x_0$ and $\gamma(1) = x$.

13.11 Example (**Planar Pendulum**) The Hamiltonian of the planar pendulum is, after normalizing its length and the acceleration of gravity, equal to

$$H(p_\psi, \psi) = \tfrac{1}{2} p_\psi^2 - \cos(\psi)\,,$$

just as in the Example 8.11 of the bead on a circular wire that is *not* rotating. Hereby, ψ denotes the angle with respect to the lower equilibrium. As can be read off the Taylor expansion $-\cos(\psi) = -1 + \frac{1}{2}\psi^2 + \mathcal{O}(\psi^4)$, the linearized expansion at the lower equilibrium is as for the harmonic oscillator with frequency 1.

The change in the frequency as the total energy $h \geq -1$ changes can be found from the formula for the time derivative of the angle: $\frac{\mathrm{d}\psi}{\mathrm{d}t} = p_\psi = \sqrt{2(h + \cos(\psi))}$. Therefore, the pendulum takes the time

$$t_h(\psi) = \frac{1}{\sqrt{2}} \int_0^{\psi} \bigl(h + \cos(x)\bigr)^{-1/2}\, \mathrm{d}x \tag{13.3.13}$$

to reach the angle ψ. According to the value $h = 1$ for the potential energy at the upper equilibrium $(p_\psi, \psi) = (0, \pi)$, we have to distinguish three cases:

- $\boldsymbol{h = 1}$: The energy surface $H^{-1}(h)$ consists of the upper equilibrium and two more orbits, called *homoclinic*, because they connect this saddle point with itself. Sometimes, $H^{-1}(h)$ is called *separatrix*, because it is the boundary separating phase space domains with qualitatively different behavior.
  On the homoclinic orbits, the time parameter equals ± 1 times

$$t_1(\psi) = \tfrac{1}{2} \int_0^{\psi} \bigl(\cos(x/2)\bigr)^{-1}\, \mathrm{d}x = \int_0^{z} (1 - y^2)^{-1}\, \mathrm{d}y\,,$$

  with the substitution $z = \sin(\psi/2)$. Therefore $t_1(\psi) = \frac{1}{2}\log(\frac{1+z}{1-z})$, i.e., the approach to the unstable upper equilibrium $\psi = \pm\pi$ occurs in a time that diverges logarithmically with the difference in angle.
- $\boldsymbol{h \in (-1, 1)}$: In this case, the energy surface consists of a single orbit. The total energy is not sufficient to reach the upper equilibrium, and the maximal angle is

of absolute value $\psi_h := \arccos(-h)$. Plugging this into (13.3.13), $t_h(\psi)$ equals

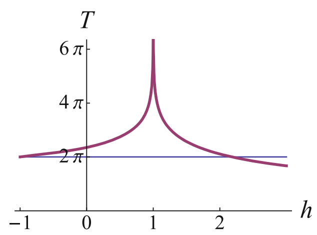
$$\frac{1}{\sqrt{2}}\int_0^{\psi}\big(\cos(x)-\cos(\psi_h)\big)^{-1/2}\,\mathrm{d}x = \tfrac{1}{2}\int_0^{\psi}\big(\sin^2(\psi_h/2)-\sin^2(x/2)\big)^{-1/2}\,\mathrm{d}x\,.$$

Reference to a table of integrals, or a computer algebra system, and solving for the angle, yields,

$$\sin(\psi/2) = \sin(\psi_h/2)\,\mathrm{sn}\big(t\,;\,\sin(\psi_h/2)\big)\,,$$

where sn is one of Jacobi's elliptic functions, called *sinus amplitudinis*. So the period of the pendulum is $T(h) := 4K(\sin(\psi_h/2))$, where K is the *complete elliptic integral of the first kind*,

$$K(k) := \int_0^{\pi/2}\big(1-k^2\sin(\varphi)^2\big)^{-1/2}\,\mathrm{d}\varphi\,.$$



Period of the planar pendulum

It can be seen from this formula that the period increases monotonically from 2π to ∞ as the energy increases from -1 to 1.
The angle φ (in the sense of action-angle variables on the invariant torus) is proportional to the time t, with constant of proportionality $2\pi/T(h)$.

- $\boldsymbol{h \in (1,\infty)}$: In this case, the energy suffices to rotate in constant direction, and accordingly, the energy surface has two connected components. One obtains analogously

$$\sin(\psi/2) = \mathrm{sn}\Big(\sqrt{(1+h)/2}\,t\,;\,\sqrt{2/(1+h)}\Big)\,,$$

and the period $T(h) = 2\sqrt{2/(1+h)}\,K\big(\sqrt{2/(1+h)}\big)$ now decreases monotonically to zero in the limit $h \nearrow +\infty$. The angle variable φ is defined just as in the previous case. ◊

13.12 Exercise (**Action of the Planar Pendulum**) Determine the action I as a function of the energy h by calculating the area of the domain $\{(p_\psi,\psi)\in\mathbb{R}\times S^1 \mid H(p_\psi,\psi)\le h\}$, in terms of complete elliptic integrals. ◊

13.4 The Momentum Mapping

Under favorable circumstances, if the Hamiltonian has symmetries, one can calculate constants of motion from these, ideally even solve the equations of motion.

Classically, such a connection between symmetries and conserved quantities is made by *Noether's theorem*. Nowadays, one frequently uses the *momentum mapping*, which is also discussed in this chapter.

The Example of Motion in a Centrally Symmetric Potential

The notion of integrability given in Definition 13.2 is often too narrow for this purpose. We can see this in Example 13.1 of planar motion in a centrally symmetric potential.

- For one thing, in this example we just pulled the angular momentum function L, which Poisson-commutes with the Hamiltonian H, out of the hat. But it is important to see how such phase space functions can be *calculated* if the symmetry of H is known.
- On the other hand, it was already assumed as known that the particle moves within a plane in the configuration space $\mathbb{R}^3_q$; this can admittedly be *proved*. To do this however, one needs to admit constants of motion that do not commute.

We begin with the second observation, thus considering the motion of a particle of mass m in $\mathbb{R}^3_q$ under the influence of a *central force* with potential $V : \mathbb{R}^3_q \to \mathbb{R}$ (that is $V(O\,q) = V(q)$ for $O \in \mathrm{O}(3)$) and the Hamiltonian on the phase space $P := \mathbb{R}^3_p \times \mathbb{R}^3_q$,

$$H : P \to \mathbb{R} \quad , \quad H(p,q) := \frac{\|p\|^2}{2m} + V(q)\,.$$

The components of the angular momentum vector are the phase space functions

$$L = \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} : P \to \mathbb{R}^3 \quad , \quad L(p,q) = q \times p = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \times \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} q_2p_3 - q_3p_2 \\ q_3p_1 - q_1p_3 \\ q_1p_2 - q_2p_1 \end{pmatrix}. \tag{13.4.1}$$

One has

$$\{L_1, L_2\} = \sum_{i=1}^{3} \left(-\frac{\partial L_1}{\partial p_i}\frac{\partial L_2}{\partial q_i} + \frac{\partial L_2}{\partial p_i}\frac{\partial L_1}{\partial q_i} \right) = -q_2p_1 + (-p_2)(-q_1) = L_3\,,$$

$$\{L_2, L_3\} = L_1\;, \quad \text{and} \quad \{L_3, L_1\} = L_2\,.$$

On the other hand, by the radial symmetry of $V : \mathbb{R}^3 \to \mathbb{R}$, there exists a function $W : [0,\infty) \to \mathbb{R}$ with $V(q) \equiv W(\|q\|)$. This implies

$$\begin{aligned}\{L_1, H\} &= \frac{1}{2m}\{L_1, p_1^2 + p_2^2 + p_3^2\} + \{L_1, W(\|q\|)\} \\ &= \frac{1}{2m}\sum_{i=1}^{3} \frac{\partial(q_2p_3 - q_3p_2)}{\partial q_i}\frac{\partial(p_1^2+p_2^2+p_3^2)}{\partial p_i} - \sum_{i=1}^{3} \frac{\partial(q_2p_3 - q_3p_2)}{\partial p_i}\frac{\partial\|q\|}{\partial q_i}\mathrm{D}W(\|q\|)\end{aligned}$$

$= 0$. Analogously, one also has $\{L_2, H\} = \{L_3, H\} = 0$.

By rotation $(p, q) \mapsto (Op, Oq)$ with a matrix $O \in \mathrm{SO}(3)$, we can always achieve that $L_1(p, q) = L_2(p, q) = 0$. Then p and q have to lie in the 1–2 plane, and we can transition to the new phase space $\mathbb{R}^2_p \times \mathbb{R}^2_q \subset \mathbb{R}^3_p \times \mathbb{R}^3_q$ and the restricted Hamiltonian on it,

$$H\restriction_{\mathbb{R}^2_p \times \mathbb{R}^2_q} : \mathbb{R}^2_p \times \mathbb{R}^2_q \to \mathbb{R} \quad , \quad (p, q) \mapsto \frac{p_1^2 + p_2^2}{2m} + W\left(\sqrt{q_1^2 + q_2^2}\right).$$

The renamed third component of the angular momentum

$$L : \mathbb{R}^2_p \times \mathbb{R}^2_q \to \mathbb{R} \quad , \quad L(p, q) = q_1 p_2 - q_2 p_1$$

remains a constant of motion in this restriction process. In this case, we have even three constants of motion for the motion defined by H, namely the components of the angular momentum; together with H, they form a system of four independent functions. But the components of the angular momentum are not mutually in involution. So the method of integration from the preceding chapter does not apply. So what is the *systematic method* underlying the reduction to the 1–2 plane that we just used?

Symplectic Actions of Groups

Let us start by making the notion of symmetry more precise. We first assume that there is a symmetry group G acting on the symplectic manifold (P, ω) that leaves the Hamiltonian $H : P \to \mathbb{R}$ invariant. Thus for the smooth action of a Lie group G with Lie algebra $\mathfrak{g}$ (see Appendix E),

$$\Phi : G \times P \to P \quad \text{and} \quad \Phi_g(p) := \Phi(g, p)\,, \tag{13.4.2}$$

we assume that

$$H \circ \Phi_g = H \qquad (g \in G). \tag{13.4.3}$$

Only such group actions are useful that are compatible with the symplectic structure of the phase space. At this point, it is not clear what is meant by this compatibility. Three alternatives come to mind naturally:

13.13 Definition

- *The group action (13.4.2) is called* **symplectic** *if the diffeomorphisms $\Phi_g : P \to P$ are symplectomorphisms, i.e., if*

$$\Phi_g^* \omega = \omega \qquad (g \in G).$$

- *A symplectic action of a group is called* **weakly Hamiltonian** *if the vector fields $X_\xi : P \to TP \quad (\xi \in \mathfrak{g})$ are Hamiltonian (i.e., the 1-form $\mathbf{i}_{X_\xi}\omega$ is exact).*
- *It is called* **Hamiltonian** *if there exists a linear mapping*

$$F : \mathfrak{g} \to C^\infty(P, \mathbb{R}) \tag{13.4.4}$$

that is a homomorphism of the Lie algebras $\mathfrak{g}$ *and* $(C^\infty(P, \mathbb{R}), \{\cdot, \cdot\})$ *(i.e., the* $\mathbb{R}$*-vector space of phase space functions with Poisson bracket (10.2.1)) and that satisfies*[7]*:*

$$X_\xi = X_{F(\xi)} \qquad (\xi \in \mathfrak{g}). \tag{13.4.5}$$

Our goals are, to the extent possible,

1. to find a simple test whether a symplectic action of a group is Hamiltonian, and then to calculate a function F satisfying (13.4.4);
2. to show that for a symmetry (13.4.3) of H, this function F is a constant of the motion under the Hamiltonian H (i.e., $dF(\xi)(X_H) = 0 \quad (\xi \in \mathfrak{g})$),
3. and to reduce the dimension of the phase space P by means of it.

For a group action Φ, the following implications apply:

$$\Phi \text{ Hamiltonian} \quad \Longrightarrow \quad \Phi \text{ weakly Hamiltonian} \quad \Longrightarrow \quad \Phi \text{ symplectic.}$$

But the converse implications do not hold[8]:

13.14 Examples (**Symplectic Group Actions**)

1. (**symplectic, but not weakly Hamiltonian**)
   We already know from Example 10.10 that on the 2-torus $P := \mathbb{T}^2$ with the volume form ω as a symplectic form, the constant vector fields are locally Hamiltonian. In this example, the action of the group $\mathbb{R}$ is the conditionally periodic motion in the direction of the vector field, and it is symplectic.
   But among them, only the vector field that is constant zero is Hamiltonian, because only then is the cohomology class $[\mathbf{i}_v\omega] \in H^1(\mathbb{T}^2) \cong \mathbb{R}^2$ equal to 0, see Example B.54.
2. (**weakly Hamiltonian, but not Hamiltonian**)
   On the phase space $P := \mathbb{R}^2_x$ with canonical symplectic form $\omega_0 = dx_1 \wedge dx_2$, the action of the group $G := \mathbb{R}^2$, given by the translations

   $$\Phi_g(x) := x + g \quad (x \in P,\ g \in G),$$

   is weakly Hamiltonian, with the Hamilton function

   $$F : \mathfrak{g} \to C^\infty(P, \mathbb{R})\ ,\ F(\xi)(x) := \langle x, \mathbb{J}\xi \rangle = x_2\xi_1 - x_1\xi_2 \quad (\xi \in \mathfrak{g} = \mathbb{R}^2, x \in P)$$

   for the matrix $\mathbb{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. But the Poisson bracket

---

[7]Here the vector field X_ξ is the infinitesimal generator of ξ introduced in Definition E.32, whereas $X_{F(\xi)}$ is the Hamiltonian vector field of the function $F(\xi) : P \to \mathbb{R}$.

[8]There are obstructions that lie in the cohomology group $H^1(P, \mathbb{R})$ and in a cohomology group $H^2(\mathfrak{g}, \mathbb{R})$ called Lie algebra cohomology, see McDuff and Salamon [MS].

$$\{F(\xi), F(\eta)\} = \frac{\partial F(\xi)}{\partial x_1}\frac{\partial F(\eta)}{\partial x_2} - \frac{\partial F(\xi)}{\partial x_2}\frac{\partial F(\eta)}{\partial x_1} = \langle \eta, \mathbb{J}\xi\rangle$$

is different from 0 when ξ, η are linearly independent.
The linear mapping $F : \mathfrak{g} \to C^\infty(P)$ is not a homomorphism of Lie algebras, because the Lie algebra $\mathfrak{g}$ is commutative (i.e., $[\xi, \eta] = 0$).
While we can still add constants to the Hamilton functions $F(\xi)$ without affecting the vector fields $X_{F(\xi)}$, this does still not change the Poisson brackets.

3. (**Hamiltonian**)
The action of the group $\mathbb{R}^k$ that is generated on a symplectic manifold (P, ω) by k Poisson-commuting functions $F_1, \ldots, F_k \in C^\infty(P, \mathbb{R})$ is Hamiltonian (of course, this is under the assumption that the flows generated by F_i exist). One sets $F : \mathbb{R}^k \to C^\infty(P, \mathbb{R})$, $F(\xi) := \sum_{i=1}^k \xi_i F_i$. In particular, an integrable system (Definition 13.2) generates a Hamiltonian action of this group. ♢

In the following, we will only study Hamiltonian actions of groups.

If we assume that the phase space P is connected, then two mappings $F^{(1)}, F^{(2)} : \mathfrak{g} \to C^\infty(P, \mathbb{R})$ in (13.4.4) that generate the same vector fields only differ by a linear mapping $F^{(1)} - F^{(2)} : \mathfrak{g} \to \mathbb{R}$, namely an element of the dual Lie algebra $\mathfrak{g}^*$.

Momentum Mappings for Lifted Configuration Space Symmetries

13.15 Example (**Rotation Group**)
$G := \mathrm{SO}(3)$ acts on the phase space $P := T^*\mathbb{R}^3 \cong \mathbb{R}^3_p \times \mathbb{R}^3_q$ by

$$\Phi_g(p, q) := \big(g(p), g(q)\big) \qquad \big(g \in \mathrm{SO}(3),\ (p, q) \in P\big), \tag{13.4.6}$$

and we will see shortly that this is a Hamiltonian action. First we calculate a Hamilton function $F : \mathfrak{g} \to C^\infty(P, \mathbb{R})$ that is parametrized by the Lie algebra $\mathfrak{g} = \mathfrak{so}(3)$ of G. Thus $\mathfrak{so}(3)$ is the three dimensional real vector space of antisymmetric 3×3 matrices $\xi \in \mathfrak{so}(3)$, and it acts by the vector fields

$$X_\xi : P \to TP \quad , \quad X_\xi(p, q) = (\xi\, p, \xi\, q) \qquad \big(\xi \in \mathfrak{so}(3)\big).$$

They are Hamiltonian vector fields for the Hamilton functions

$$F(\xi) : P \to \mathbb{R} \quad , \quad F(\xi)(p, q) = \langle p, \xi\, q\rangle \qquad \big(\xi \in \mathfrak{so}(3)\big). \tag{13.4.7}$$

The isomorphism

$$i : \mathbb{R}^3 \to \mathfrak{so}(3) \quad , \quad \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \mapsto \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix} \tag{13.4.8}$$

with the property $i(a)b = a \times b \quad (a, b \in \mathbb{R}^3)$, which was already useful in relation (6.3.15) for magnetic fields, leads to

$$F \circ i(a) = \langle a, L \rangle \tag{13.4.9}$$

with the angular momentum vector $L = \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix} : P \to \mathbb{R}^3$ from the example on page 344.

So in particular for $a \in S^2$, the function $F \circ i(a)$ is the angular momentum in direction a, and this Hamilton function generates a right turn of period 2π about the axis $\mathrm{span}(a)$ oriented in the direction of a. ◊

In the example, the group action Φ on the phase space T^*M arose from a group action on the configuration space M. Namely, Φ_g was the cotangent lift $T_g^* : T^*M \to T^*M$ of a diffeomorphism $g^{-1} : M \to M$ (see Definition 10.32). Such examples occur more frequently; they lead to Hamiltonian actions Φ and allow for a simple calculation of F (and hence a partial answer to the first question on page 346):

13.16 Theorem (**Cotangent Lift**)

1. *For a group action* $\Psi : G \times M \to M$, *the* **left lift**

$$\Psi^L : G \times P \to P \quad , \quad \Psi_g^L = T^*\Psi_{g^{-1}}$$

to $P := T^*M$ *is also a group action.*[9]

2. *The left lift is a symplectic action of the group* G *and even leaves the tautological 1-form* θ_0 *on* P *(see Definition 10.7) invariant.*
3. *If a symplectic action* $\Phi : G \times P \to P$ *of the group* G *on* P *leaves the tautological form* θ_0 *on* $P = T^*M$ *invariant, then it is a Hamiltonian action, and (13.4.4) is of the form*

$$F : \mathfrak{g} \to C^\infty(P, \mathbb{R}) \quad , \quad F(\xi) := \mathbf{i}_{X_\xi}\theta_0 \, .$$

Proof: We use that $\omega_0 = -d\theta_0$.

- Ψ_g^L maps the cotangent space T_q^*M to $T_{\Psi_g(q)}^*M$, and statement 1 follows from (10.3.2): $\Psi_{g\circ h}^L = T^*\Psi_{(g\circ h)^{-1}}$ equals

$$T^*\Psi_{h^{-1}\circ g^{-1}} = T^*(\Psi_{h^{-1}} \circ \Psi_{g^{-1}}) = T^*(\Psi_{g^{-1}}) \circ T^*(\Psi_{h^{-1}}) = \Psi_g^L \circ \Psi_h^L \, .$$

- By Theorem 10.35, the cotangent lift is exact symplectic. Hence statement 2.
- For all $\xi \in \mathfrak{g}$, the function $F(\xi)$ is a Hamilton function for X_ξ, because invariance implies that $L_{X_\xi}\theta_0 = 0$, and therefore

$$dF(\xi) = d\mathbf{i}_{X_\xi}\theta_0 = L_{X_\xi}\theta_0 - \mathbf{i}_{X_\xi}d\theta_0 = 0 + \mathbf{i}_{X_\xi}\omega \, .$$

- In order to calculate the Poisson bracket of $F(\xi)$ and $F(\eta)$ for $\xi, \eta \in \mathfrak{g}$, we write

[9] More precisely a left action. If we were to use Ψ_g instead of $\Psi_{g^{-1}}$, we would get a right action.

$$\begin{aligned}\big\{F(\xi), F(\eta)\big\} &= L_{X_\xi} F(\eta) = L_{X_\xi} \mathbf{i}_{X_\eta} \theta_0 \\ &= (L_{X_\xi} \mathbf{i}_{X_\eta} - \mathbf{i}_{X_\eta} L_{X_\xi})\theta_0 = \mathbf{i}_{[X_\xi, X_\eta]}\theta_0 = \mathbf{i}_{X_{[\xi,\eta]}}\theta_0 = F\big([\xi,\eta]\big).\end{aligned}$$

Here we have used Lemma 10.24 in the fourth equality. We have thus shown that $\xi \mapsto F(\xi)$ is a Lie algebra homomorphism. □

13.17 Example (Rotation Group)
The group actions Φ_g from Example 13.15 are cotangent lifts of the *inverse* rotations $g^{-1} : \mathbb{R}^3 \to \mathbb{R}^3$, with $g \in \mathrm{SO}(3)$, because $\big(g^{-1}\big)^\top = g$.

Also, the mapping $F : \mathfrak{so}(3) \to C^\infty(P, \mathbb{R})$ defined by (13.4.7) is defined as required by Theorem 13.16. ◊

Therefore, whenever a Lie group G acts on a manifold M by diffeomorphisms, we obtain a Hamiltonian action Φ of G on the cotangent bundle T^*M.

The more commonly used wording for Hamiltonian symmetries is dual to the one given here and accordingly uses the dual $\mathfrak{g}^*$ of the Lie algebra $\mathfrak{g}$:

13.18 Definition

- *A mapping $J : P \to \mathfrak{g}^*$ is called* **momentum mapping** *for the symplectic action Φ of the group G, if the linear mapping*

$$\hat{J} : \mathfrak{g} \to C^\infty(P) \quad , \quad \big(\hat{J}(\xi)\big)(x) := J(x)(\xi) \qquad (\xi \in \mathfrak{g},\ x \in P)$$

  induced by J satisfies the analog of (13.4.5), namely
  $X_\xi = X_{\hat{J}(\xi)}$.
- *Let G be a group and M, N sets with group actions $\Phi : G \times M \to M$ and $\Psi : G \times N \to N$. A mapping $f : M \to N$ is called G-***equivariant***, if*

$$\Psi_g \circ f = f \circ \Phi_g \quad (g \in G),$$

$$\begin{array}{ccc} M & \xrightarrow{\Phi_g} & M \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ N & \xrightarrow{\Psi_g} & N \end{array}$$

  i.e., if the diagram to the right commutes.

Equivariant means: varying in the same way. Applied to the momentum mapping, on the phase space P we have the action Φ of the Lie group G, whereas G acts on the dual Lie algebra $\mathfrak{g}^*$ by the coadjoint representation defined in (E.4.2).

So in the case of a momentum mapping, which will then be called Ad^*-*equivariant*, the diagram to the right will commute.

$$\begin{array}{ccc} P & \xrightarrow{\Phi_g} & P \\ {\scriptstyle J}\downarrow & & \downarrow{\scriptstyle J} \\ \mathfrak{g}^* & \xrightarrow{\mathrm{Ad}^*_{g^{-1}}} & \mathfrak{g}^* \end{array}$$

13.19 Example (Rotation Group)
The angular momentum $L : P \to \mathbb{R}^3$ in (13.4.1) is a momentum mapping and as such, properly speaking, a mapping with values in $\mathfrak{so}(3)^*$. Accordingly, the $\mathbb{R}^3$ scalar product in (13.4.9) is actually the pairing of $\mathfrak{so}(3)$ and $\mathfrak{so}(3)^*$.

Since $(O\,a) \times (O\,b) = O\,(a \times b)$ for all rotation matrices[10] $O \in \mathrm{SO}(3)$ and all $a, b \in \mathbb{R}^3$, the Ad^*-equivariance follows. ◊

Among the two equivalent conditions in the following lemma, the second is local in nature (because it only refers to the Lie algebra $\mathfrak{g}$), but the first is global (because it refers to the Lie group G). The Lie group needs to be connected so that one can conclude from the local to the global property.

13.20 Lemma *A symplectic group action $\Phi : G \times P \to P$ of a connected Lie group G is Hamiltonian if and only if there exists an* Ad^**-equivariant momentum mapping $J : P \to \mathfrak{g}^*$ for Φ.*

Proof:

- Let $J : P \to \mathfrak{g}^*$ be an Ad^*-equivariant momentum mapping for Φ. It is to be shown that $\hat{J}$ from Definition 13.18 is Hamiltonian.
  The first condition $X_\xi = X_{\hat{J}(\xi)}$ is satisfied by definition. It remains to be shown that the Poisson bracket satisfies the equation

$$\{\hat{J}(\xi), \hat{J}(\eta)\} = \hat{J}\big([\xi, \eta]\big) \qquad (\xi, \eta \in \mathfrak{g}).$$

  To this end, we use the relation

$$\{\hat{J}(\xi), \hat{J}(\eta)\} = -L_{X_{\hat{J}(\xi)}} \hat{J}(\eta) = -L_{X_\xi} \hat{J}(\eta)$$

  between Poisson bracket and Lie derivative (Theorem 10.16). By Theorem B.34, using the exponential map $\exp : \mathfrak{g} \to G$ from (E.3.1), it is true for any $x \in P$ that

$$\begin{aligned}\big(L_{X_\xi} \hat{J}(\eta)\big)(x) &= \frac{\mathrm{d}}{\mathrm{d}t} \hat{J}(\eta)\big(\Phi(\exp(t\xi), x)\big)\Big|_{t=0} = \left\langle \frac{\mathrm{d}}{\mathrm{d}t} J\big(\Phi(\exp(t\xi), x)\big)\Big|_{t=0}, \eta \right\rangle \\ &= \left\langle \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ad}^*_{\exp(-t\xi)}\Big|_{t=0} J(x), \eta \right\rangle = \left\langle J(x), \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ad}_{\exp(-t\xi)}\Big|_{t=0} \eta \right\rangle \\ &= \big\langle J(x), [-\xi, \eta] \big\rangle = -\hat{J}\big([\xi, \eta]\big)(x)\,,\end{aligned}$$

  where we have used the Ad^*-equivariance of J and (E.4.3).

- Conversely, let Φ be Hamiltonian for $F : \mathfrak{g} \to C^\infty(P, \mathbb{R})$. We claim that $J : P \to \mathfrak{g}^*,\ J(x)(\xi) := F(\xi)(x)$ is an Ad^*-equivariant momentum mapping. Only the Ad^*-equivariance is to be shown yet, namely

$$J\big(\Phi_g(x)\big) = \mathrm{Ad}^*_{g^{-1}}\big(J(x)\big) \qquad (g \in G). \tag{13.4.10}$$

As this is true for the identity $g = e$, and G is connected, it suffices to show that the derivatives of both sides of (13.4.10) with respect to g are equal.

[10] But *not* for the orthogonal matrices $O \in \mathrm{O}(3) \setminus \mathrm{SO}(3)$!

Since $\Phi_g \circ \Phi_h = \Phi_{gh}$ by the defining property of group actions, it suffices to prove this equality at $g = e$. As the image $\exp(\mathfrak{g}) \subseteq G$ of the exponential map (E.3.1) is a neighborhood of $e \in G$, that claim is equivalent to showing for all $\xi \in \mathfrak{g}$ that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[J\big(\Phi_{\exp(t\xi)}(x)\big) - \mathrm{Ad}^*_{\exp(-t\xi)}\big(J(x)\big)\right]\Big|_{t=0} = 0\,.$$

This in turn in equivalent to the claim

$$L_{X_\xi} F(\eta) = -F\big([\xi,\eta]\big) \qquad (\eta \in \mathfrak{g}),$$

which is a consequence of $L_{X_\xi} F(\eta) = -\{F(\xi), F(\eta)\}$ and the homomorphism property of F. □

Now we can easily achieve the second of our goals stated on page 346:

13.21 Lemma *For a Hamiltonian action Φ of a group, with F from (13.4.4), the Φ-invariance (13.4.3) of a Hamilton function $H : P \to \mathbb{R}$, implies that F is a constant of the motion by generated H.*

Proof: It is to be shown for all $\xi \in \mathfrak{g}$, that $\{H, F(\xi)\} = 0$. By (13.4.3), this is however a consequence of $\{H, F(\xi)\} = L_{X_{F(\xi)}} H = \frac{\mathrm{d}}{\mathrm{d}t} H \circ \Phi_{\exp(t\xi)}|_{t=0} = 0$. □

We have thus proved the theorem by EMMY NOETHER [No] that continuous symmetries generate constants of motion:

13.22 Theorem (**Noether**)
If a Hamiltonian $H : P \to \mathbb{R}$ on the symplectic phase space (P, ω) generates the flow Ψ and is invariant under a Hamiltonian action $\Phi : G \times P \to P$ of a group G, then there exists a momentum mapping $J : P \to \mathfrak{g}^*$ of Φ with

$$J \circ \Psi_t = J \qquad (t \in \mathbb{R}).$$

13.5 * Reduction of the Phase Space

"The more I have learned about physics, the more convinced I am that physics provides, in a sense, the deepest applications of mathematics. The mathematical problems that have been solved, or techniques that have arisen out of physics in the past, have been the lifeblood of mathematics... The really deep questions are still in the physical sciences. For the health of mathematics at its research level, I think it is very important to maintain that link as much as possible." (MICHAEL ATIYAH, in: Mathematical Intelligencer, **6**, 9–19 (1984)

Symplectic Reduction

One one hand, the Noether theorem tells us that in the presence of continuous symmetries, there exist constants of motion, which then allow us (for a regular value $j \in \mathfrak{g}^*$

of the momentum mapping) to study the motion on the submanifold $M_j := J^{-1}(j)$ of P.

But this submanifold will in general not be a symplectic manifold any more (as for instance in Example 13.15, where the momentum mapping is the angular momentum and M_j ist 3-dimensional, so it cannot have a symplectic form at all).

As a matter of fact, one can often return to a Hamiltonian system by means of a further reduction of the phase space dimension, called the *symplectic* or *Marsden-Weinstein reduction* [MW]. This we also have already seen in the case of a centrally symmetric potential (Example 13.1).

But this new phase space P_j arises from M_j by taking a quotient, rather than being a submanifold; the points in P_j are the orbits of the action of a subgroup G_j of G. Despite the quotient space construction, P_j still has the structure of a manifold, and the quotient map $M_j \to P_j$ is a principal bundle (see Definition F.4).

We now discuss this reduction technique. All mappings that occur in this chapter are smooth.

13.23 Theorem (**Marsden and Weinstein**)
For the Ad^**-equivariant momentum mapping* $J : P \to \mathfrak{g}^*$ *of the symplectic action* $\Phi : G \times P \to P$ *of the group* G*, let the point* j *in the image be a regular value of* J *(so that* $M_j := J^{-1}(j)$ *is a submanifold, with inclusion* $i_j : M_j \to P$*).*

Assume that the isotropy group $G_j := \{g \in G \mid \mathrm{Ad}^*_g(j) = j\}$ *acts freely and properly on* M_j*. Then*

$$\pi_j : M_j \to P_j := M_j/G_j \tag{13.5.1}$$

is a submersion onto a manifold P_j*, and* P_j *has a unique symplectic structure* ω_j *with the property*

$$\boxed{i_j^*\omega = \pi_j^*\omega_j \ .} \tag{13.5.2}$$

13.24 Remarks (**Generalizations**)

1. It is only for ascertaining the regularity of (13.5.1) that we use that the action of G_j on M_j is free and proper.
2. Another case of interest could be if the action of G_j is locally free.[11] In this case, P_j may not be a manifold, but will be a kind of object called *orbifold* (see Audin, Cannas da Silva and Lerman [ACL]). G_j always acts locally freely if j is a regular value of J. ♢

Proof of Theorem 13.23:
By Theorem A.46 about regular values, M_j is a submanifold of P.

- On this submanifold, the isotropy group G_j acts because of the Ad^*-equivariance of J.
- By Theorem E.36, P_j is a manifold, and $\pi_j : M_j \to P_j$ is a surjective submersion.

[11] **Definition**: A topological group action $\Phi : G \times M \to M$ is *locally free* if there exists a neighborhood $U \subseteq G$ of e such that the following implication holds: If $\Phi_g(x) = x$ for some $x \in M$ and some $g \in U$, then $g = e$.

- For this reason, pullbacks of different differential forms by π_j^* are also different. So there can be at most one symplectic form ω_j on P_j that satisfies (13.5.2).
- As π_j is a surjective submersion, we can define the 2-form ω_j on P_j by

$$\omega_j\big(T_m\pi_j(v)\,,\;T_m\pi_j(w)\big) := \omega(v,w) \qquad (v,w\in T_mM_j)\,,$$

  independent of a choice of representatives if one has the condition:

$$\omega(v',w') = \omega(v,w) \text{ if } T_{m'}\pi_j(v') = T_m\pi_j(v) \text{ and } T_{m'}\pi_j(w') = T_m\pi_j(w)$$

  for all points $m' := \Phi_g(m) \in M_j$ on the G_j-orbit and all tangent vectors $v', w' \in T_{m'}M_j$.
- If $m' = \Phi_g(m)$, then $\Phi_g^*\omega = \omega$, so we can assume $m = m'$ without loss of generality.
- To this end it suffices to show that

$$\omega(h,k) = 0 \qquad (k\in T_mM_j) \tag{13.5.3}$$

  for all tangent vectors $h \in \ker(T_m\pi_j)$. But there exists a vector $\xi \in \mathfrak{g}$ for which the infinitesimal generator $X_\xi = \frac{\mathrm{d}}{\mathrm{d}t}\Phi_{\exp(t\xi)}|_{t=0}$ at m equals h (so $X_\xi(m) = h$). Now $\mathbf{i}_{X_\xi}\omega = dJ(\xi)$ is a 1-form that vanishes on M_j, because $J(\xi) : P \to \mathbb{R}$ is constant on $M_j = J^{-1}(j)$. Equation (13.5.3) is thus proved.
- The 2-form ω_j on P_j that we have just defined is closed. For on one hand, (B.25) and (13.5.2) imply

$$\pi_j^* d\omega_j = d\pi_j^*\omega_j = di_j^*\omega = i_j^* d\omega = 0\,.$$

  On the other hand, since π_j is a surjective submersion, one can conclude $\varphi = 0$ from $\pi_j^*\varphi = 0$ for any differential form φ on P_j.
- Also, ω_j is not degenerate and thus a symplectic form on P_j. We see this because for $m \in M_j,\ v \in T_mP$, and $\xi \in \mathfrak{g}$, one has

$$\omega\big(X_\xi(m), v\big) = d\hat{J}(\xi)(v) = \langle T_mJ(v), \xi\rangle\,.$$

  This expression vanishes for all $\xi \in \mathfrak{g}$ if and only if $T_mJ(v) = 0$, i.e., if and only if v lies in the subspace T_mM_j of T_mP.
  In other words, the ω-orthogonal complement of T_mM_j equals the subspace $U := \{X_\xi(m) \mid \xi \in \mathfrak{g}\}$ of T_mP.
- It suffices to show that the Lie algebra $\mathrm{Lie}(G_j)$ of the isotropy group G_j satisfies

$$U \cap T_mM_j = \{X_\xi(m) \mid \xi \in \mathrm{Lie}(G_j)\}\,,$$

  because this is the kernel of $T_m\pi_j : T_mM_j \to T_{\pi_j(m)}P_j$. But $X_\xi(m)$ lies in T_mM_j, if and only if $dJ(X_\xi)(m) = 0$, or by the Ad^*-equivariance of J, if $j \in \mathfrak{g}^*$ is a fixed point of the ad-operator ad^*_ξ (see (E.4.4)). This is the case if and only if $\xi \in \mathrm{Lie}(G_j)$. □

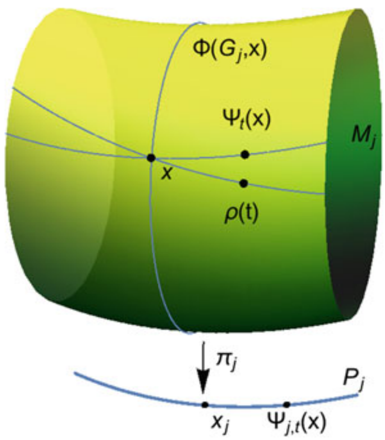
As will soon become apparent in examples, this symplectic, or Marsden-Weinstein reduction has applications outside the area of dynamical systems. It is in particular a method for finding new symplectic manifolds.

Nevertheless, we will first have a look at the case where the action Φ is used to simplify Hamiltonian differential equations.

13.25 Theorem *Under the hypotheses of Theorem 13.23, let the Hamilton function $H : P \to \mathbb{R}$ be invariant under the group action Φ and generate a flow $\Psi : \mathbb{R} \times P \to P$.*

1. *The function $H_j : P_j \to \mathbb{R}$ defined by the relation*

$$i_j^* H = \pi_j^* H_j \qquad (13.5.4)$$



generates a Hamiltonian flow Ψ_j on the symplectic manifold (P_j, ω_j), and this flow is a factor of the restriction of the flow Ψ to M_j.

2. *If for the initial value $x \in M_j$ and $x_j := \pi_j(x) \in P_j$, the curve $t \mapsto \rho(t) \in M_j$ is a lift of the solution $t \mapsto \Psi_{j,t}(x_j) \in P_j$ with $\rho(0) = x$, then the solution $t \mapsto \Psi_t(x) \in M_j$ can be represented in the form $\Psi_t(x) = \Phi_{g(t)}\big(\rho(t)\big)$ (see the figure).*
The curve $t \mapsto g(t) \in G_j$ in the isotropy group satisfies (with L_g denoting the left action from (E.1.3)) the initial value problem

$$g(0) = e \quad , \quad g'(t) = T_e L_{g(t)} \xi(t) \quad \textit{with} \quad X_{\xi(t)}\big(\rho(t)\big) := X_H\big(\rho(t)\big) - \rho'(t)\,. \qquad (13.5.5)$$

Proof:

1. • To begin with, the Hamilton function H is invariant under the symmetry Φ, so Equation (13.5.4) does have a solution $H_j : P_j \to \mathbb{R}$. This solution is unique because π_j is a surjective submersion.

   • The flow Ψ can be restricted to the submanifold $M_j = J^{-1}(j)$ because of Noether's theorem (Theorem 13.22). It follows from (13.5.2) and (13.5.4) that the Hamiltonian vector field X_H on M_j is projected to the Hamiltonian vector field X_{H_j} on P_j by the linearized bundle projection $T\pi_j$. So the flows have the factor property

$$\pi_j \circ \Psi_t(x) = \Psi_{j,t} \circ \pi_j(x) \qquad (x \in M_j,\ t \in \mathbb{R}).$$

2. • Again, we first need to check the solvability of an equation, namely this time

$$\xi(t) \in \mathrm{Lie}(G_j) \quad \text{with} \quad X_{\xi(t)}\big(\rho(t)\big) = X_H\big(\rho(t)\big) - \rho'(t) \in T_{\rho(t)} M_j\,.$$

The existence of such a $\xi(t) \in \mathrm{Lie}(G_j)$ is due to the fact that (by the lift property of ρ) the vector $X_H\big(\rho(t)\big) - \rho'(t) \in T_{\rho(t)} M_j$ lies in the kernel of the linearized projection $T_{\rho(t)}\pi_j : T_{\rho(t)} M_j \to T_{\psi_{j,t}(x)} P_j$.
This also determines $\xi(t) \in \mathrm{Lie}(G_j)$ uniquely, because by assumption, the isotropy group G_j acts freely on M_j.

• To verify (13.5.5), we next derive the equality $\Psi_t(x) = \Phi\big(g(t), \rho(t)\big)$ with respect to time. Its left hand side yields $\frac{\mathrm{d}}{\mathrm{d}t}\Psi_t(x) = X_H\big(\Psi_t(x)\big)$, hence

$$\big(T_{\rho(t)}\Phi_{g(t)}\big)^{-1} X_H\big(\Psi_t(x)\big) = X_H\big(\rho(t)\big)\,. \tag{13.5.6}$$

• The time derivative of the right hand side is

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi\big(g(t), \rho(t)\big) = T_{\rho(t)}\Phi_{g(t)}\rho'(t) + T_{g(t)}\tilde{\Phi}_{\rho(t)} g'(t)\,, \tag{13.5.7}$$

with

$$\tilde{\Phi}_x : G \to P \quad , \quad \tilde{\Phi}_x(g) = \Phi(g, x)\,.$$

We transform the second term in (13.5.7) as follows. From the definition of group actions, we obtain the relation

$$\tilde{\Phi}_x = \Phi_g \circ \tilde{\Phi}_x \circ L_{g^{-1}} \qquad (g \in G,\ x \in P).$$

Therefore,

$$T_{g(t)}\tilde{\Phi}_{\rho(t)} = \big(T_{\rho(t)}\Phi_{g(t)}\big) \circ \big(T_e\tilde{\Phi}_{\rho(t)}\big) \circ \big(T_{g(t)}L_{g^{-1}(t)}\big)\,. \tag{13.5.8}$$

Thus with $g'(t) = T_e L_{g(t)}\xi(t)$ and $T_e\tilde{\Phi}_{\rho(t)}\xi = X_\xi\big(\rho(t)\big)$ $(\xi \in \mathfrak{g})$ and (13.5.8), one obtains (13.5.7) in the form

$$\big(T_{\rho(t)}\Phi_{g(t)}\big)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}\Phi\big(g(t), \rho(t)\big) = \rho'(t) + X_{\xi(t)}\big(\rho(t)\big)\,. \tag{13.5.9}$$

(13.5.6) and (13.5.9) display the relation that defines $\xi(t)$ in (13.5.5). □

Applications of the Symplectic Reduction

"manche meinen lechts und rinks kann man nicht velwechsern werch ein illtum!" lichtung,[12] by Ernst Jandl

By Theorem 13.25, the integration of the differential equation on P is reduced to the integration of a differential equation on the reduced phase space P_j and, as we will next see, of a differential equation on the isotopy group G_j.

13.26 Examples (**Reduction**)

1. As a compact Lie group, the k-dimensional torus $G = \mathbb{T}^k$ always acts properly, and the isotropy group G_j coincides with G. Therefore, for regular values $j \in \mathfrak{g} \cong \mathbb{R}^k$ of the momentum mapping, one has

$$\dim(P_j) = \dim(P) - 2k\,.$$

In the extreme case of an integrable system with $k = n := \frac{1}{2}\dim(P)$, the dimension of the reduced phase space P_j is zero. In this case, the compact connected components $\tilde{M}_j$ of M_j are themselves diffeomorphic to the torus $\mathbb{T}^n$. The task of integrating the Φ-invariant vector field $X_H\restriction_{\tilde{M}_j}$ on the fibers $\tilde{M}_j$ of the bundle $M_j \to P_j$ was solved in Chapter 13.2.
However, it must be noted that it is initially the group $\mathbb{R}^n$ that acts on P. It is a preprocessing step prior to the Marsden-Weinstein reduction that identifies the periods (and thus the torus action).

2. Let us look one last time at the already overused example of the action (13.4.6) by the rotation group SO(3) on the phase space $P := T^*\mathbb{R}^3$. The momentum mapping (13.4.1) of the angular momentum $L : P \to \mathfrak{so}(3)$ has $0 \in \mathfrak{so}(3)$ as its only singular value. So for values $\ell \neq 0$ of the angular momentum, the pre-images $M_\ell := L^{-1}(\ell)$ are manifolds. With the 2-dimensional subspace $\ell^\perp := \{q \in \mathbb{R}^3 \mid \langle q, \ell\rangle = 0\}$, they are then of the form

$$M_\ell = \left\{ \left(\tfrac{\ell\times q}{\|q\|^2} + cq\,,\, q \right) \,\middle|\, c \in \mathbb{R},\; q \in \ell^\perp\backslash\{0\} \right\} \subset P\,. \tag{13.5.10}$$

This can be checked by plugging the expression in (13.5.10) into the definition $L(p, q) = q \times p$ of the angular momentum.
By parametrization with (c, q), one obtains the diffeomorphism

$$M_\ell \cong \mathbb{R} \times \left(\mathbb{R}^2\backslash\{0\}\right). \tag{13.5.11}$$

The isotropy group $\mathrm{SO}(3)_\ell$ consists of the lifted rotations (13.4.6) about the axis $\mathrm{span}(\ell)$. As this rotation group, which is isomorphic to SO(2), is compact, it always acts properly. It also acts freely, namely only on the second factor

[12] Translator's comment: With due aporogies to the Engrish leadels, this phonetic pun poetly is an impossibirity to tlansrate; best shot (but not as concise) is: "Some leckon and berieve that light and reft cannot be confused: What an ellol, what a farracy!".

in (13.5.11), so that $(\mathbb{R}^2\backslash\{0\})/\mathrm{SO}(2) \cong \mathbb{R}^+$. This can be seen by use of polar coordinates in $\mathbb{R}^2\backslash\{0\}$. The hypotheses of Theorem 13.23 are thus satisfied.
One obtains the reduced phase space $P_\ell = M_\ell/\mathrm{SO}(3)_\ell \cong \mathbb{R} \times \mathbb{R}^+$, which was already used in Example 13.1 to analyze the radial motion.

3. The Hamiltonian $H : P \to \mathbb{R}$, $H(p,q) = \frac{1}{2}\|p\|^2$ generates the *free motion* $\Phi_t(p,q) = (p, q + tp)$ on $P := T^*\mathbb{R}^d$. This Hamiltonian action by the group $\mathbb{R}$ is free and proper, if restricted to an energy surface $\Sigma_E = H^{-1}(E)$ for energy $E > 0$.
For instance when $E := \frac{1}{2}$, one has $\Sigma_E \cong S_p^{d-1} \times \mathbb{R}_q^d$, and the reduced phase space is diffeomorphic to
$$P_E := \Sigma_E/\mathbb{R} \cong T^* S^{d-1}, \tag{13.5.12}$$
because for each point $(p,q) \in \Sigma_E$, there is exactly one time $t \in \mathbb{R}$ for which $\langle p, q + tp\rangle = 0$.
This cotangent bundle P_E (or the tangent bundle $T S^{d-1}$ respectively) was already used in the context of scattering in a potential, see (12.1.16). By Corollary 12.8, the action of the Hamiltonian flow in the potential is also proper and free for high energies E. While we cannot necessarily parametrize the reduced phase space (13.5.12) in the same way as in the case without potential, we can parametrize it by the asymptotic scattering data. ◊

A mathematically important class of examples is given by the coadjoint action of a Lie group G on its dual Lie algebra $\mathfrak{g}^*$, see (E.4.2).

13.27 Example (**Coadjoint Action**)
For the Lie group SO(3), one has $\mathfrak{g} = \mathrm{Alt}(3,\mathbb{R})$. We identify $\mathfrak{g}^*$ with $\mathrm{Alt}(3,\mathbb{R})$ as well, by writing the pairing of Lie algebra and dual Lie algebra in the form of a trace:
$$\big\langle \xi^*, \eta\big\rangle = \mathrm{tr}(\xi^*\eta) \qquad \big(\xi^*, \eta \in \mathrm{Alt}(3,\mathbb{R})\big).$$
Then from $\mathrm{Ad}_g(\eta) = g\eta g^{-1}$, one obtains the coadjoint representation in the form $\mathrm{Ad}^*_{g^{-1}}(\xi^*) = g\xi^* g^{-1}$. By Exercise E.31, with its identification $i : \mathbb{R}^3 \to \mathfrak{so}(3)$ from (13.4.8), the adjoint representation is given by $\mathrm{Ad}_g(\eta) = g\,\eta$, hence the dual formula
$$\mathrm{Ad}^*_{g^{-1}}(\xi^*) = g\,\xi^* \qquad \big(g \in \mathrm{SO}(3),\ \xi^* \in \mathbb{R}^3\big).$$
In this sense, the SO(3)-orbit of $\xi^* \in \mathbb{R}^3 \setminus \{0\}$ is equal to the 2-sphere
$$\{g\,\xi^* \mid g \in \mathrm{SO}(3)\} = \{x \in \mathbb{R}^3 \mid \|x\| = \|\xi^*\|\}$$
of radius $\|\xi^*\|$, and the origin is an orbit itself. ◊

In this example, the orbits have symplectic forms that are invariant under the coadjoint action, namely the area forms. But it has not become apparent yet how these forms would have arisen out of the symplectic reduction. Making this connection in all generality is due to Kirillov, Kostant and Souriau.

13.28 Theorem *The orbits of the coadjoint action of a Lie group have a natural symplectic form that is invariant under this action.*

Proof: The idea is to obtain these orbits by Marsden-Weinstein reduction.

- We begin by observing that the cotangent bundle $P := T^*G$ of the Lie group G is a symplectic manifold, just like any cotangent bundle (P, ω_0) with $\omega_0 = -d\theta_0$ and the tautological form θ_0 from Definition 10.7.

- So by Theorem 10.35, the left action

$$L_g : G \to G \quad , \quad h \mapsto g \circ h \qquad (g \in G)$$

of G can be lifted to a symplectic action

$$\Phi_g := T^* L_g : P \to P \qquad (g \in G).$$

By Theorem 13.16, this action has the momentum mapping

$$J : P \to \mathfrak{g}^* \quad , \quad J(\alpha_g)(\xi) = \alpha_g\big(X_\xi(g)\big) \qquad \big(\xi \in \mathfrak{g},\ g \in G\big), \tag{13.5.13}$$

where $X_\xi : G \to TG$ is the infinitesimal generator for ξ, i.e., the *right* invariant vector field with $X_\xi(e) = \xi$. So we can write $X_\xi(g) \in T_g G$ in the form $X_\xi(g) = T_e R_g(\xi)$, or dually, $J(\alpha_g)(\xi) = (T_e R_g)^* \alpha_g(\xi)$.

- Thus for $j \in \mathfrak{g}^*$, the level set of the (Ad^*-equivariant) momentum mapping is

$$M_j = J^{-1}(j) = \mathrm{graph}(\alpha_j) \subset T^*G\,,$$

where $\alpha_j \in \Omega^1(G)$ is the right invariant 1-form whose value in $e \in G$ is $\alpha_j(e) = j$. Therefore, M_j is diffeomorphic to G. Since the group action $\Phi : G \times P \to P$ is a left action, the isotropy group

$$G_j = \{g \in G \mid \mathrm{Ad}_g^*(j) = j\}$$

for the Ad^*-action by G equals the subgroup $\{g \in G \mid L_g^*(\alpha_j) = \alpha_j\}$ of those *left* translations that do not change the *right* invariant 1-form. Thus the orbit $\{\mathrm{Ad}_g^*(j) \mid g \in G\} \subseteq \mathfrak{g}^*$ of j is of the form $G/G_j \cong P_j$ with the reduced symplectic phase space (P_j, ω_j) from Theorem 13.23. □

In general, the coadjoint orbits are only immersed, not imbedded, submanifolds of $\mathfrak{g}^*$, see Example 14.1.6 of MARSDEN and RATIU [MR]. In any case, being symplectic manifolds, they are of even dimension.

13.29 Exercise (Coadjoint Orbits)
Show for the case of a subgroup $G \le \mathrm{GL}(n, \mathbb{R})$, and for the orbit $\mathcal{O}(\xi) \subset \mathfrak{g}^*$ of the coadjoint action of a Lie group G on its dual Lie algebra $\mathfrak{g}^*$,

(a) that the tangent space $T_\xi \mathcal{O}(\xi)$ consists of vectors of the form $\mathrm{ad}_u^* \xi \quad (u \in \mathfrak{g})$ (where the operator ad_u^* on $\mathfrak{g}^*$ is adjoint to the operator ad_u in (E.4.4),
(b) that the symplectic structure on $\mathcal{O}(\xi)$ is of the form

$$\omega_\xi\left(\mathrm{ad}_u^*\xi,\ \mathrm{ad}_u^*\xi\right) = -\langle \xi, [u, v]\rangle \qquad (u, v \in \mathfrak{g}), \tag{13.5.14}$$

(c) that the momentum mapping $J : \mathcal{O}(\xi) \to \mathfrak{g}^*$ is the inclusion of the orbit. ◊

13.30 Example (Special Linear Group)
For $n = 1$, the Lie group $\mathrm{Sp}(2n, \mathbb{R})$, which is of paramount importance in classical mechanics, coincides with the special linear group $\mathrm{SL}(2, \mathbb{R}) = \{M \in \mathrm{Mat}(2, \mathbb{R}) \mid \det(M) = 1\}$ (Exercise 6.26). As it is a matrix group, its Lie algebra is

$$\mathfrak{sl}(2, \mathbb{R}) = \{\xi \in \mathrm{Mat}(2, \mathbb{R}) \mid \mathrm{tr}(\xi) = 0\}.$$

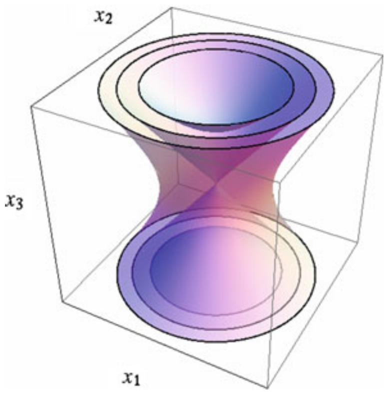
Again we identify the dual Lie algebra $\mathfrak{sl}(2, \mathbb{R})^*$ with $\mathfrak{sl}(2, \mathbb{R})$ by means of the trace and parametrize it by

$$i : \mathbb{R}^3 \to \mathfrak{sl}(2, \mathbb{R})^* \quad , \quad x \mapsto \begin{pmatrix} x_1 & x_2+x_3 \\ x_2-x_3 & -x_1 \end{pmatrix}.$$

Thus

$$\det\left(i(x)\right) = x_3^2 - x_1^2 - x_2^2,$$

and this quantity is invariant under the coadjoint action. The level sets are therefore quadrics in $\mathbb{R}^3$, namely, depending on the sign, rotational hyperboloids of one or of two sheets, or in the case of vanishing determinant, a double cone (see figure). Each connected component of a quadric makes one orbit, with the exception of the double cone. This latter consists of three orbits, namely the origin and the two simple cones without their vertex (corresponding to the orbits of the infinitesimally symplectic matrices $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$).



Orbits in $\mathfrak{sl}(2, \mathbb{R})^*$

To these orbits, the exponential map assigns orbits of symplectic matrices with the same eigenvalues. Compare therefore the family of quadrics with a neighborhood of the identity in the representation of the symplectic group $\mathrm{Sp}(2, \mathbb{R})$ given on page 97. ◊

13.31 Remarks (**Coadjoint Actions**)

1. Theorem 13.28 is not an isolated result, but rather is the starting point of a method by KIRILLOV called *orbit method* (which he presents in his survey article [Ki]). In this method, symplectic geometry is chosen as the approach to the representation theory of Lie groups. The corresponding theory in physics is called *geometric quantization*.
2. While the dual Lie algebras studied in Theorem 13.28 cannot be viewed as symplectic vector spaces in any natural way (they often have odd dimension, as is the case in the example $\mathfrak{so}(3)$), they do have a Poisson structure.
   A *Poisson structure* on a manifold P is a bilinear mapping
   $$\{\cdot,\cdot\} : C^\infty(P,\mathbb{R}) \times C^\infty(P,\mathbb{R}) \to C^\infty(P,\mathbb{R}),$$
   that makes $C^\infty(P,\mathbb{R})$ into a Lie algebra and satisfies the derivation property:
   $$\{fg,h\} = f\{g,h\} + \{f,h\}g \qquad \big(f,g,h \in C^\infty(P,\mathbb{R})\big).$$
   This generalizes the notion of Poisson bracket of a symplectic manifold from Def. 10.15. In the case of dual Lie algebras, the orbits arise as something called *symplectic leaves* of a Poisson structure (see MARSDEN and RATIU [MR]). ◊

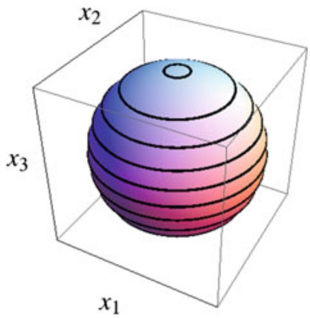
Reduction of Symplectic Torus Actions

We now study torus actions in particular.

13.32 Exercise (**Hamiltonian S^1-Action**)
Show that the height function

$$H : S^2 \to \mathbb{R},\ H(x) = x_3$$

on the 2-sphere $S^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$ (with the area form as its symplectic form) gives rise to an S^1-action, namely the rotation about the 3-axis, see the figure.
Compare also with Exercise 10.30. ◊



In this case, the image of the symplectic manifold under the momentum mapping is $H(S^2)$, i.e., an interval.
We generalize this example of a torus action by viewing S^2 as $\mathbb{CP}(1)$ and looking for analogous actions on the projective space $\mathbb{CP}(d)$.

13.33 Example (**Hamiltonian $\mathbb{T}^d$-Action on $\mathbb{CP}(d)$**)
In dimension $n \in \mathbb{N}$, the abelian group $\mathbb{T}^n := \{t = (t_1,\ldots,t_n) \in \mathbb{C}^n \mid |t_k| = 1\}$ acts by pointwise multiplication

$$\tilde{\Phi} : \mathbb{T}^n \times \mathbb{C}^n \to \mathbb{C} \quad , \quad \Phi(t,x) = \big(t_1x_1,\ldots,t_nx_n\big)$$

on the symplectic vector space $(\mathbb{C}^n, \omega)$ with $\omega(u, v) = \mathrm{Im}(\langle u, v\rangle)$ (see Remark 6.15.1). This action is Hamiltonian, with an Ad^*-equivariant (hence invariant) momentum mapping

$$\tilde{J} : \mathbb{C}^n \to \mathbb{R}^n \cong \mathrm{Lie}(\mathbb{T}^n)^* \quad , \quad x \mapsto -\tfrac{1}{2}\big(|x_1|^2, \dots, |x_n|^2\big)\,.$$

On the other hand, $S^1 \subset \mathbb{C}$ also has a Hamiltonian action on $\mathbb{C}^n$ by

$$\Psi : S^1 \times \mathbb{C}^n \to \mathbb{C}^n \quad , \quad \Psi(s, x) = sx\,,$$

with Hamilton function $H(x) = -\frac{1}{2}\|x\|^2$. This function has the same form as for the harmonic oscillators with equal frequencies as discussed in Theorem 6.35. So the reduced phase space M_h/S^1 (for $h = -\frac{1}{2}$, hence $M_h = H^{-1}(h) \cong S^{2n-1}$) is the complex projective space $\mathbb{C}\mathrm{P}(n-1)$.

The group actions Ψ and Φ commute, so for all $t \in \mathbb{T}^n$ and $x \in \mathbb{C}^n$, orbits will be mapped to orbits: $\tilde{\Phi}_t\big(\mathcal{O}_\Psi(x)\big) = \mathcal{O}_\Psi\big(\tilde{\Phi}_t(x)\big)$. Thus we obtain an action by $\mathbb{T}^n$ on the symplectic phase space (P, ω), where P is the projective space $\mathbb{C}\mathrm{P}(d)$ with the real dimension $2d$ (for $d := n-1$).

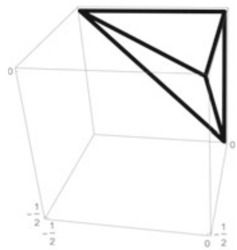
However, this action is not free any more, because for $s \in S^1$ and $\tilde{s} := (s, \dots, s) \in \mathbb{T}^n$, one has $\tilde{\Phi}_{\tilde{s}} = \Psi_s$. But if we consider the subgroup

$$G \;:=\; \big\{(t_1, \dots, t_n) \in \mathbb{T}^n \mid t_n = 1\big\} \;\cong\; \mathbb{T}^d$$

of $\mathbb{T}^n$, then $\mathbb{T}^n \cong G \oplus S^1$, and we obtain a free Hamiltonian action $\Phi : G \times P \to P$. Its momentum mapping is of the form

$$J : P \to \mathbb{R}^d \cong \mathrm{Lie}(G) \quad , \quad J([x]) = -\tfrac{1}{2}\left(\frac{|x_1|^2}{\|x\|^2}, \dots, \frac{|x_d|^2}{\|x\|^2}\right).$$

Here the notation is to be understood in such a way that we represent points $[x] \in \mathbb{C}P(d)$ by their representatives $x \in \mathbb{C}^{d+1}\backslash\{0\}$, where $[x] = [y]$ if there exists $\lambda \in \mathbb{C}\backslash\{0\}$ such that $x = \lambda y$.
Thus the image of the momentum mapping is the simplex $\Delta_d := J(P) \subset \mathbb{R}^d$ with the vertices 0 and $-\frac{1}{2}e_k,\ k = 1, \dots, d$. ◊
As in Example 13.1 (see Figure 13.1.1 left), we obtain a polyhedron (namely the intersection of finitely many half spaces) as the image of a momentum mapping for a torus action.



Simplex Δ_3

This is not a coincidence. M. Atiyah in [At], and V. Guillemin and S. Sternberg in [GS2] proved the following theorem independently:

13.34 Theorem *If $\Phi : \mathbb{T}^m \times P \to P$ is a Hamiltonian action of a torus on the compact and connected symplectic manifold (P, ω) with momentum mapping $J : P \to \mathbb{R}^m$, then:*

1. *The level sets $M_j = J^{-1}(j) \subset P$ are connected.*
2. *The image $J(P) \subset \mathbb{R}^m$ is a polytope, specifically the convex hull of the images $J(p) \in \mathbb{R}^m$ of the fixed points $p \in P$ of the Φ-action.*

13.35 Literature A proof of this theorem can also be found in Chapter 5.4 of McDuff and Salamon [MS].

It is precisely known which polytopes can occur as images of the momentum mapping. These polytopes, called *Delzant polytopes* are described for instance in the article by Ana Cannas da Silva in [ACL], and they are important in algebraic topology and string theory. ◊

The Theorem by Schur and Horn

From a knowledge of the eigenvalues λ_k of a Hermitian matrix $A = A^* \in \mathrm{Herm}(d, \mathbb{C})$, we can obtain the trace of A as the sum of these eigenvalues. In order to uniquely assign to such a matrix a vector of eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$, we assume the eigenvalues are numbered in order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$, and thus we get a mapping

$$\Lambda : \mathrm{Herm}(d, \mathbb{C}) \to \mathbb{R}^d \,. \qquad (13.5.15)$$

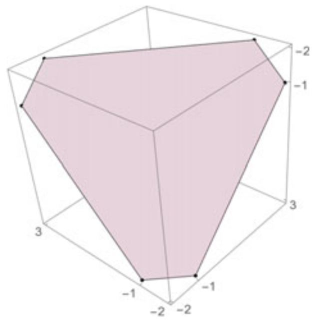
We are now interested in the diagonals of the isospectral matrices from $\Lambda^{-1}(\lambda)$, i.e., we consider the projection

$$\Pi : \mathrm{Herm}(d, \mathbb{C}) \to \mathbb{R}^d,\ (a_{i,k})_{i,k=1}^d \mapsto (a_{k,k})_{k=1}^d .$$

As diagonal matrices have their diagonal entries as eigenvalues, it is clear that the set

$$\Pi\big(\Lambda^{-1}(\lambda)\big) \subset \mathbb{R}^d \qquad (13.5.16)$$

contains the points $\lambda_\sigma := (\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(d)})$ for every permutation $\sigma \in S_d$.



The polytope $\Pi\big(\Lambda^{-1}(\lambda)\big)$ for the eigenvalues $3, -1, -2$

The theorem by Schur and Horn states now that this set arises by convex combinations from these diagonal matrices:

13.36 Theorem (**Schur and Horn**)
For the Hermitian matrices with eigenvalues λ, one has the equality

$$\Pi\big(\Lambda^{-1}(\lambda)\big) = \mathrm{conv}\big(\{\lambda_\sigma \mid \sigma \in S_d\}\big) \qquad (13.5.17)$$

Proof: Nowadays[13] the proof can be given by applying Theorem 13.34 (see ATIYAH [At]). One can figure it out by asking which torus action can be used:

- To begin with, $\mathrm{Herm}(d,\mathbb{C})$ is a Lie algebra with the Lie bracket $[A,B] := i(AB - BA)$. But as we are seeking a compact Lie group for the Lie algebra, it is better to transition to antisymmetric matrices. These make up the Lie algebra $\mathfrak{u}(d)$ of the unitary group $\mathrm{U}(d)$, and $\mathfrak{u}(d)$ is isomorphic to $\mathrm{Herm}(d,\mathbb{C})$ (by multiplication with i).
- If we again identify the Lie algebra with its dual (via $A \mapsto \mathrm{tr}(A\,\cdot)$), then $\mathrm{U}(d)$ acts on $\mathfrak{u}(d)^*$, and thus also on $\mathrm{Herm}(d,\mathbb{C})$, by the coadjoint mapping.
- As multiplication with unitary matrices corresponds to those changes of the basis that preserve the scalar product, the $\mathrm{U}(d)$-orbit

$$\mathcal{O}_{\mathrm{U}(d)}(A) = \left\{UAU^{-1} \mid U \in \mathrm{U}(d)\right\} \qquad \big(A \in \mathrm{Herm}(d,\mathbb{C})\big)$$

  consists precisely of all Hermitian matrices with the same eigenvalues $\lambda \in \mathbb{R}^d$ as A. Therefore $\mathcal{O}_{\mathrm{U}(d)}(A) = \Lambda^{-1}(\lambda)$, and this manifold has, by Theorem 13.28, a symplectic structure that is invariant under the $\mathrm{U}(d)$-action.
- We first consider the generic (that is, typical) case where the eigenvalues are pairwise different, ($\lambda_{k+1} < \lambda_k$, $k = 1,\ldots,d-1$). The isotropy subgroup of diagonal matrices A from $\Lambda^{-1}(\lambda)$ is then of the form $\mathbb{T}^d \subset \mathrm{U}(d)$, i.e., it consists of all those diagonal matrices whose diagonal entries are complex numbers of modulus 1. On one hand, this implies[14]

$$\Lambda^{-1}(\lambda) \cong \mathrm{U}(d)/\mathbb{T}^d\,.$$

  On the other hand, the subgroup $\mathbb{T}^d$ acts on $\Lambda^{-1}(\lambda)$ (nontrivially if $d \geq 2$) by

$$\Phi : \mathbb{T}^d \times \Lambda^{-1}(\lambda) \to \Lambda^{-1}(\lambda) \quad , \quad \Phi_U(A) = UAU^{-1}\,,$$

  and the only fixed points of this action are the diagonal matrices $\mathrm{diag}(\lambda_\sigma) \in \Lambda^{-1}(\lambda)$ corresponding to the permutations λ_σ.
- By Exercise 13.29.c, the momentum mapping of Φ is of the form

$$J : \Lambda^{-1}(\lambda) \to \mathfrak{t}^* \quad , \quad A \mapsto \mathrm{tr}(A\,\cdot) = \mathrm{tr}\big(\Pi(A)\,\cdot\big)\,,$$

  with the projection (13.5.16) and the abelian Lie algebra $\mathfrak{t} = \mathrm{Lie}(\mathbb{T}^d)$ of real diagonal matrices, which is isomorphic to $\mathbb{R}^d$.
- Thus, in the generic case, the conclusion follows from Theorem 13.34. The case of degenerate eigenvalues follows by passing to the limit. □

[13]The inclusion '⊆' of the identity (13.5.17) was proved by I. Schur in 1924, the converse inclusion by A. Horn in 1954, which is long before the symplectic reduction in the version of Theorem 13.23 was known.

[14]thus for instance $\Lambda^{-1}(\lambda) \cong S^2$ for $d = 2$, because $\mathrm{U}(2)/S^1 \cong \mathrm{SU}(2) \cong S^3$ and $\mathrm{SU}(2)/S^1 \cong S^2$, see the Hopf map on page 117.

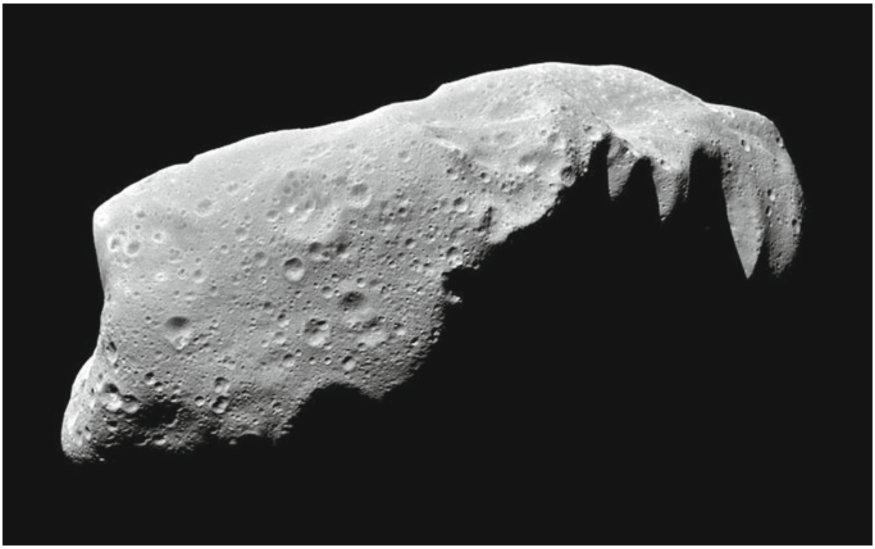
13.37 Exercise (Ky-Fan Maximum Principle for Hermitian Matrices) Conclude from the Schur-Horn theorem that the eigenvalues $(\lambda_1, \ldots, \lambda_d) = \Lambda(A)$ of $A \in \mathrm{Herm}(d, \mathbb{C})$ satisfy the linear relations

$$\sum_{i=1}^{k} \lambda_i = \max_{(x_1,\ldots,x_k)\in V_k(\mathbb{C}^d)} \sum_{i=1}^{k} \langle x_i, Ax_i\rangle \qquad (k = 1, \ldots, d).$$

Here $V_k(\mathbb{C}^d)$ denotes the compact *Stiefel manifold* of orthonormal k-tuples $(x_1, \ldots, x_k)$ of vectors $x_i \in \mathbb{C}^d$; the real dimension of this manifold is $(2dk - k^2)$. ♢

13.38 Literature A modification of the Schur-Horn theorem describes the structure of eigenvalues of the matrix $A + B$ under the assumption that only the eigenvalues of A and $B \in \mathrm{Herm}(d, \mathbb{C})$ are known. This theorem can be proved by a generalization of Theorem 13.34 that goes back to Frances Kirwan, see for instance the article [Li] by P. Littelmann. ♢

Chapter 14
Rigid and Non-Rigid Bodies



The asteroid Ida[1] rotates with a period of 4.6 hours about its axis of maximum moment of inertia. The sun causes a precession of this axis with a period of 77 000 years. Picture: courtesy of NASA/JPL-Caltech.

Until now, we have mostly studied the motion of point masses. They are a good idealization for many natural phenomena, like for instance in celestial mechanics. But often, we have to deal with extended solids. Sometimes, as in the case of liquids, these can only be described by means of continuum mechanics, i.e., with partial rather than ordinary differential equations. In the case of *rigid bodies* however, the

[1]Ida has an average diameter of about 30 Km and a moon that orbits Ida with pedestrian speed. The picture was taken in 1993 by the satellite Galileo.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_14

distances between atoms (viewed as points) remain constant in time, and this allows a description by ordinary differential equations.

14.1 Motions of Euclidean Space

We begin with Euclidean space $(\mathbb{R}^d, \langle\cdot,\cdot\rangle)$ and its canonical scalar product $\langle a, b\rangle = \sum_{k=1}^d a_k b_k$, norm $\|a\| = \sqrt{\langle a, a\rangle}$, and metric $d(a,b) = \|a-b\|$, as well as its *isometries* or *motions*, i.e., distance preserving mappings $I : \mathbb{R}^d \to \mathbb{R}^d$.

We recall a fact known from linear algebra. Obviously, every mapping $I : \mathbb{R}^d \to \mathbb{R}^d,\ I(q) = Oq + v$ with $O \in \mathrm{O}(d)$ (the orthogonal group consisting of rotations and rotoreflections of $\mathbb{R}^d$, see Appendix E) and $v \in \mathbb{R}^d$ is an isometry, because $\|I(a) - I(b)\| = \|O(a-b)\| = \|a-b\|$. Conversely, one has:

14.1 Theorem (Euclidean Motions) *For every isometry $I : \mathbb{R}^d \to \mathbb{R}^d$, there exist unique $O \in \mathrm{O}(d)$ and $v \in \mathbb{R}^d$ such that $I(q) = Oq + v \quad (q \in \mathbb{R}^d)$.*

Proof: Let $I : \mathbb{R}^d \to \mathbb{R}^d$ be an isometry.

- We set $v := I(0)$; then $\tilde{I}(q) := I(q) - v$ leaves the origin invariant.
- From the polarization identity for the scalar product

$$\langle a, b\rangle = \tfrac{1}{4}\big(\|a+b\|^2 - \|a-b\|^2\big) \qquad \big(a, b \in \mathbb{R}^d\big)$$

  and

$$\|\tilde{I}(c)\| = d\big(\tilde{I}(0), \tilde{I}(c)\big) = d(0,c) = \|c\| \qquad \big(c \in \mathbb{R}^d\big)\ ,$$

  it follows that $\left\langle \tilde{I}(a), \tilde{I}(b)\right\rangle = \langle a, b\rangle$. If $\tilde{I}$ is linear, then $\tilde{I}$ is orthogonal.
- The linearity of $\tilde{I}$ follows from the relation

$$\left\|\tilde{I}(\alpha a + \beta b) - \alpha\tilde{I}(a) - \beta\tilde{I}(b)\right\|^2 = 0 \qquad \big(\alpha, \beta \in \mathbb{R},\ a, b \in \mathbb{R}^d\big)\,.$$

  Indeed, the left side is a sum of scalar products, and therefore, due to the invariance of scalar products under $\tilde{I}$, it is equal to $\|(\alpha a + \beta b) - \alpha a - \beta b\|^2 = 0$. Therefore $\tilde{I}$ is linear. □

Isometries of metric spaces are always injective, but not necessarily surjective (find a counterexample!). In the case of $\mathbb{R}^d$ however, this is the case, and we obtain the group of isometries of $\mathbb{R}^d$, the *Euclidean group* $\mathbb{E}(d)$.

- The Euclidean group is a Lie group (see Appendix E), and by Theorem 14.1, as a manifold, it equals

$$\mathbb{E}(d) = \mathbb{R}^d \times \mathrm{O}(d)\,. \tag{14.1.1}$$

- The group structure of $\mathbb{E}(d)$, however, is not that of a direct product of the two groups, but is rather their *semidirect product*, a notion explained in Appendix E.1. So the composition of two isometries is given by

$$(v_2, O_2) \circ (v_1, O_1) = \big(O_2 v_1 + v_2,\ O_2 O_1\big) \qquad \big(v_i \in \mathbb{R}^d,\ O_i \in \mathrm{O}(d)\big).$$

14.2 Exercise Which is the inverse element to $(v, O) \in \mathbb{E}(d)$? ◊

The subgroup of *orientation preserving* isometries of $\mathbb{R}^d$ is denoted as

$$\mathbb{SE}(d) := \{(v, O) \in \mathbb{E}(d) \mid O \in \mathrm{SO}(d)\}\,. \tag{14.1.2}$$

14.2 Kinematics of Rigid Bodies

In a rigid body consisting of n mass points, the distances $d_{k,\ell}$ between the kth and the ℓth mass point remain constant in time.

On the other hand, we cannot simply define a rigid body in $\mathbb{R}^d$ by prescribing $\binom{n}{2}$ positive numbers $d_{k,\ell} = d_{\ell,k},\ 1 \le k < \ell \le n$, because

- the $d_{k,\ell}$ need to satisfy certain relations like for example[2] the triangle inequality, which are necessary for points $q_1, \ldots, q_n \in \mathbb{R}^d$ with these distances to even exist.
- Even as the distances $d_{k,\ell}$ stay the same under arbitrary isometries, *physical* motions are orientation preserving.

We proceed differently and define

14.3 Definition
- *For $d, n \in \mathbb{N}$, the* **diagonal action** *of $\mathbb{E}(d)$ on $\mathbb{R}^{nd}$ is defined by*

$$\Phi^{n,d} : \mathbb{E}(d) \times \mathbb{R}^{nd} \to \mathbb{R}^{nd} \quad , \quad \Phi^{n,d}\big(I, (q_1, \ldots, q_n)\big) = \big(I(q_1), \ldots, I(q_n)\big)\,.$$

- *The $\Phi^{n,d}$-orbits $\mathcal{O} \subseteq \mathbb{R}^{nd}$ of the group $\mathbb{SE}(d)$ are called* **rigid bodies** *of n points in $\mathbb{R}^d$.*

This looks abstract, but it describes what we want to have, because $q, q' \in \mathbb{R}^{nd}$ belong to the same orbit if and only if

$$\|q'_k - q'_\ell\| = \|q_k - q_\ell\| \qquad (1 \le k < \ell \le n)\,,$$

and moreover, the orientation is preserved. The orbit $\mathcal{O}_q$ of the diagonal action of $\mathbb{SE}(d)$ through $q \in \mathbb{R}^{nd}$ has the structure of a quotient space

[2] Another condition is that certain determinants called the *Cayley-Menger determinants* of any $d+1$ points are nonnegative. An example for $d = 2$ is $-16 \det \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & d_{1,2}^2 & d_{1,3}^2 \\ 1 & d_{2,1}^2 & 0 & d_{2,3}^2 \\ 1 & d_{3,1}^2 & d_{3,2}^2 & 0 \end{pmatrix}$.

$$\mathcal{O}_q \cong \mathbb{SE}(d)/\mathbb{SE}(d)_q \,,$$

where $\mathbb{SE}(d)_q = \{I \in \mathbb{SE}(d) \mid \Phi^{n,d}(I, q) = q\}$ denotes the stabilizer group of q, see Appendix E.1.

14.4 Examples (Rigid Bodies)

1. ***n* = 1 particles**. Here the isotropy group is

$$\mathbb{SE}(d)_q = \left\{(v, O) \in \mathbb{SE}(d) \mid v = (\mathbb{1} - O)q\right\} \cong \mathrm{SO}(d) \subset \mathbb{SE}(d)\,,$$

and there is only one orbit: $\mathcal{O} \cong \mathbb{R}^d = \mathbb{SE}(d)/\mathrm{SO}(d)$.

2. ***n* = 2 particles**. If the distance is $d_{1,2} = 0$, then the orbit $\mathcal{O}_q$ through $q \in \mathbb{R}^{2d}$ is again of the form $\mathcal{O}_q = \mathbb{R}^d$. Otherwise, the isotropy group is $\mathbb{SE}(d)_q \cong \mathrm{SO}(d-1)$ (with $\mathrm{SO}(0) := \{e\}$), because we can perform arbitrary rotations in the $(d-1)$-dimensional subspace of $\mathbb{R}^d$, that is perpendicular to the line $\mathrm{span}(q_1 - q_2) \subseteq \mathbb{R}^d$. Therefore, the orbit through q is equal to

$$\mathcal{O}_q \cong \mathbb{SE}(d)/\mathbb{SE}(d)_q = \mathbb{R}^d \times \mathrm{SO}(d)/\mathrm{SO}(d-1) = \mathbb{R}^d \times S^{d-1}$$

if $d \geq 2$ (and $\mathcal{O}_q \cong \mathbb{R}$ for $d = 1$).
This can be understood directly, because for the first particle, one can choose its position q_1 freely, whereas for given distance $d_{1,2} > 0$, the second particle must lie on the sphere $\{q_2 \in \mathbb{R}^d \mid \|q_2 - q_1\| = d_{1,2}\}$.

3. ***n* ≥ *d* particles**. In an orbit $\mathcal{O} \subset \mathbb{R}^{nd}$, we can for instance freely choose the position $q_1 \in \mathbb{R}^d$ of the first particle.
If the vectors $q_2 - q_1, q_3 - q_1, \ldots, q_n - q_1 \in \mathbb{R}^d$ span $\mathbb{R}^d$, then the isotropy group $\mathbb{SE}(d)_q$ has just one element, and thus $\mathcal{O}_q \cong \mathbb{SE}(d)$.
This actually happens already if they span a $(d-1)$-dimensional subspace of $\mathbb{R}^d$, because no true rotation of $\mathbb{R}^d$ leaves the points of this subspace invariant.
For $n \geq d$ particles[3] the case $\mathcal{O}_q \cong \mathbb{SE}(d)$ is therefore typical. ◊

It is only this latter case that we will study further, and thus we obtain as rigid bodies in $\mathbb{R}^d$ just the submanifolds of $\mathbb{R}^{nd}$ that are isomorphic to $\mathbb{SE}(d)$.

The motion on this configuration space is described in the formalism of holonomic constraints from Chapter 8.2. So if the Hamilton function is of the form

$$\tilde{H} : \mathbb{R}^{nd}_p \times \mathbb{R}^{nd}_q \to \mathbb{R} \quad , \quad \tilde{H}(p, q) = \textstyle\sum_{k=1}^n \frac{\|p_k\|^2}{2m_k} + \tilde{V}(q)\,,$$

hence the Lagrange function equals

$$\tilde{\mathcal{L}} : \mathbb{R}^{nd}_q \times \mathbb{R}^{nd}_v \to \mathbb{R} \quad , \quad \tilde{\mathcal{L}}(q, v) = \sum_{k=1}^n \frac{m_k}{2}\|v_k\|^2 - \tilde{V}(q)\,,$$

[3] If for example the q_k are positions of atoms, then $d = 3$, and for a macroscopic rigid body, one may have $n = 10^{23}$.

then we parametrize the orbit $\mathcal{O}_q$, which contains the *reference configuration* $q \in \mathbb{R}^{nd}$, by the Euclidean group, i.e., we use the diffeomorphism

$$Q : \mathbb{SE}(d) \to \mathcal{O}_q \subset \mathbb{R}^{nd} \quad , \quad Q(a, O)_k := O q_k + a \quad (k = 1, \dots, n).$$

As for any Lie group, the tangent bundle of $\mathbb{SE}(d)$ is parallelizable (see Definition A.43), and using the vector space

$$\mathrm{Alt}(d, \mathbb{R}) = \{A \in \mathrm{Mat}(d, \mathbb{R}) \mid A^\top = -A\}$$

of antisymmetric matrices, it is of the form

$$T\,\mathbb{SE}(d) \cong \mathbb{SE}(d) \times \big(\mathbb{R}^d \times \mathrm{Alt}(d, \mathbb{R})\big)$$

according to (E.3.2). Thus the derivative of Q at position (a, O) can be written as

$$\mathrm{D}Q(a, O)(v, A) = (AOq_1 + v, \dots, AOq_n + v) \qquad \big(v \in \mathbb{R}^d,\ A \in \mathrm{Alt}(d, \mathbb{R})\big).$$

By definition, the Lagrange function on the tangent bundle of the orbit

$$\mathcal{L} : T\mathcal{O}_q \to \mathbb{R} \quad , \quad \mathcal{L}(x, w) = \tilde{\mathcal{L}}\big(Q(x),\ \mathrm{D}Q(x)\, w\big)$$

with $x = (a, O),\ w = (v, A)$, and $V(x) := \tilde{V}\big(Q(x)\big)$ is obtained as

$$\mathcal{L}(x, w) = \sum_{k=1}^{n} \frac{m_k}{2} \langle AOq_k + v, AOq_k + v \rangle - V(a, O)\,. \tag{14.2.1}$$

Denoting by $m_N := \sum_{k=1}^{n} m_k$ the *total mass*, we can now assume that the center of mass

$$q_N : \mathbb{R}^{nd} \to \mathbb{R}^d \quad , \quad q_N(q) = \sum_{k=1}^{n} \frac{m_k}{m_N} q_k \in \mathbb{R}^d \tag{14.2.2}$$

of the reference configuration q is at 0, because under translation of q by $\Phi^{n,d}$, one can find a point in $\mathcal{O}_q$ that has this property.

We expand the scalar product in (14.2.1). Using the orthogonal projections $P_k : \mathbb{R}^d \to \mathbb{R}^d$ onto $\mathrm{span}\big(q_k - Q\big)$ and the *tensor of inertia*[4]

$$\tilde{I} = \tilde{I}_{q_N(q)}(q) \quad \text{for} \quad \tilde{I}_Q(q) := \sum_{k=1}^{n} m_k \|q_k - Q\|^2 P_k \ \in\ \mathrm{Sym}(d, \mathbb{R}) \tag{14.2.3}$$

[4]The tilde here is meant to symbolize that in three dimensions, this definition does not coincide with the usual one (see (14.3.4)).

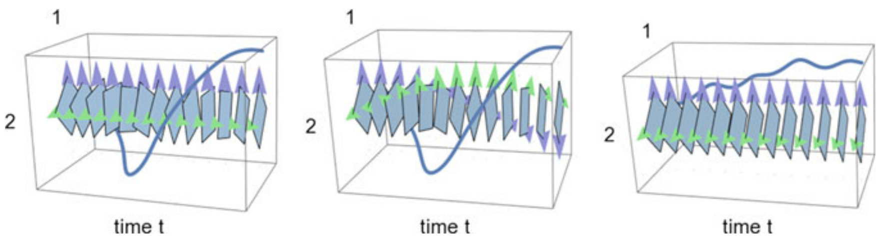


Figure 14.2.1 Moving quadrangle in extended phase space $\mathbb{R}^2 \times I$. Left: Spatially fixed basis (e_1, e_2), with graph of the curve c from Lemma 14.6, Center: Moving basis (E_1, E_2) in spatially fixed coordinates, Right: Graph of the curve C in moving basis (E_1, E_2)

with respect to $Q \in \mathbb{R}^d$, one then has

$$\mathcal{L}(x, w) = \tfrac{1}{2} m_N \|v\|^2 + \tfrac{1}{2}\mathrm{tr}\big(\tilde{I} O^{-1} A^\top A O\big) - V(a, O)\,. \tag{14.2.4}$$

In the next section, we will solve the Lagrange equations of (14.2.4) in simple cases. This will be done by means of a variety of coordinate systems.

14.5 Definition

- *The canonical basis* $(e_1, \dots, e_d)$ *of* $\mathbb{R}^d$ *is also called* **spatially fixed basis**, *and coordinates with respect to it,* **spatially fixed coordinates**.
- *For an interval* $I \subseteq \mathbb{R}$ *and a smooth curve* $O : I \to \mathrm{SO}(d)$, *the orthonormal basis*

$$\big(E_1(t), \dots, E_d(t)\big) \quad \textit{with} \quad E_k(t) := O(t)\, e_k$$

parametrized by time $t \in I$ *is called the* **moving basis** *of* $\mathbb{R}^d$. *Coordinates with respect to the moving basis are called* **moving coordinates**, *or* **body coordinates**.

With these names, we have a motion of the rigid body in mind in which at time t, the points of the body are at locations

$$q_\ell(t) := O(t)\, q_\ell(t_0) + a(t) \qquad (\ell = 1, \dots, n),$$

with $O(t_0) = 1\!\!1$ and $a(t_0) = 0$. If we use that same curve $O : I \to \mathrm{SO}(d)$ in Definition 14.5, then $\langle E_k(t), q_\ell(t)\rangle = \big\langle e_k\,,\ q_\ell(t_0) + O(t)^{-1} a(t)\big\rangle$. So if the motion has no translational part ($a(t) = 0$ $(t \in I)$), then the locations in the moving basis do not change: defining

$$Q_\ell(t) := \sum_{k=1}^{d} \langle E_k(t), q_\ell(t)\rangle\, E_k(t) \quad \text{yields} \quad Q_\ell(t) = \sum_{k=1}^{d} \langle e_k, q_\ell(t_0)\rangle\, E_k(t)\,.$$

A useful representation of these bases can be given in the *extended configuration space*

$$E := \mathbb{R}^d \times I \ \text{, with bundle projection } \pi : E \to I\ , (q,t) \mapsto t\ .$$

The spatially fixed basis in the fiber $\pi^{-1}(t)$ of the vector bundle E over the point t of the base manifold I (i.e., the time interval) will now be denoted as $\big(e_1(t), \ldots, e_d(t)\big)$. The moving basis $\big(E_1(t), \ldots, E_d(t)\big)$ is in the same fiber, rotated with respect to the spatially fixed basis (see Figure 14.2.1).

14.6 Lemma (Kinematics in Body Coordinates)
For a smooth curve $c : I \to \mathbb{R}^d$ in spatially fixed coordinates and the time dependent rotation $O : I \to \mathrm{SO}(d)$ from Definition 14.5, let $C := O^{-1}\, c : I \to \mathbb{R}^d$ denote its representation in body coordinates. Then

- *the velocity satisfies*

$$C' = O^{-1}\, c' - BC \quad \textit{with} \quad B(t) := O^{-1}(t)\, O'(t) \in \mathrm{Alt}(d,\mathbb{R}) \quad (t \in I)$$

- *and the acceleration satisfies*

$$C'' = O^{-1}\, c'' - 2BC' - B^2 C - B'C\ . \tag{14.2.5}$$

Proof:

- From the product rule $O(O^{-1})' + O'O^{-1} = (O\,O^{-1})' = 1\!\!1' = 0$ for the derivative of $O : I \to \mathrm{SO}(d)$, one gets $(O^{-1})' = -O^{-1}O'O^{-1}$. The formula for the velocity C' is obtained from $C = O^{-1}\, c$ by the product rule.

- $B(t)$ is in $\mathrm{Alt}(d,\mathbb{R})$, because $(O^{-1})' = (O^\top)' = (O')^\top$ implies

$$B^\top = (O')^\top (O^{-1})^\top = (O^{-1})'O = -(O^{-1}O'O^{-1})O = -O^{-1}O' = -B.$$

- Plugging the just obtained relation $O^{-1}\, c' = C' + BC$ into

$$\begin{aligned} C'' &= (O^{-1}\, c' - BC)' = O^{-1}\, c'' + (O^{-1})'\, c' - BC' - B'C \\ &= O^{-1}\, c'' - O^{-1}O'O^{-1}c' - BC' - B'C\ , \end{aligned}$$

one obtains the relation $C'' = O^{-1}\, c'' - O^{-1}O'C' - O^{-1}O'BC - BC' - B'C$ between the accelerations in the two coordinate systems. □

14.7 Remark (Retrograde Motion of Planets)

More generally, we can consider time dependent coordinate transformations $(a, O) : I \to \mathbb{SE}(d)$ that, at time $t \in I$, have the form $q \mapsto O(t)q + a(t)$, so they also contain af translation $a(t) \in \mathbb{R}^3$ besides the rotation $O(t) \in \mathrm{SO}(3)$.

For instance, let $a(t)$ be the position of the earth (or rather of an observer on it) at time t in a coordinate system at whose center the sun is located. We make the simplifying assumption that throughout the year, the observer always looks at the same point in the background of stars, so we set $O(t) = 1\!\!1$.

Due to the small eccentricity, the ellipse a and the corresponding ellipse of the Mars orbit may be well approximated by circles that are traversed with constant velocity. However, transforming that circle into the coordinate system of the observer of Mars leads to a complicated motion of Mars on the firmament, see Figure 14.2.2 (left and center).

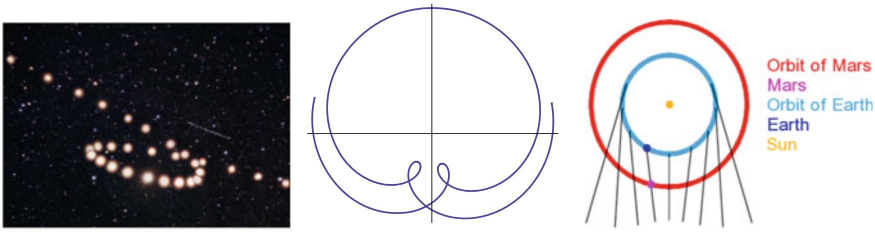


Figure 14.2.2 Left and center: Shape of the orbit of Mars as observed from the earth. Photo: Tunç Tezel. Right: Explanation of the retrograde motion

In particular, it appears to return from west to east, against its usual motion on the firmament, when the earth passes it. This apparent motion is called *retrograde*, see Figure 14.2.2 (right).

In Greek antiquity, the retrograde motion was explained by the theory of epicycles, in which the planets move along a small circle (the *epicycle*), whose center in turn moves along a circle centered at the stationary earth, called the *deferent*.

Even though this geocentric theory is kinematically equivalent to a heliocentric circular motion, due to the commutativity of vector addition, it is not equivalent dynamically. In particular, the earth undergoes no acceleration in the geocentric model. ◊

Epicycle model by Apollonius
(about 200 BC)

14.8 Exercises (Pseudoforces)

1. Show in $d = 2$ dimensions, for $\mathbb{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and a time dependent rotation matrix $O(t) = \begin{pmatrix} \cos(\varphi(t)) & -\sin(\varphi(t)) \\ \sin(\varphi(t)) & \cos(\varphi(t)) \end{pmatrix}$, that

$$C''(t) = \begin{pmatrix} \cos(\varphi(t)) & \sin(\varphi(t)) \\ -\sin(\varphi(t)) & \cos(\varphi(t)) \end{pmatrix} c''(t) - 2\varphi'(t)\mathbb{J}C'(t) + \varphi'(t)^2 C(t) - \varphi''(t)\mathbb{J}C(t).$$

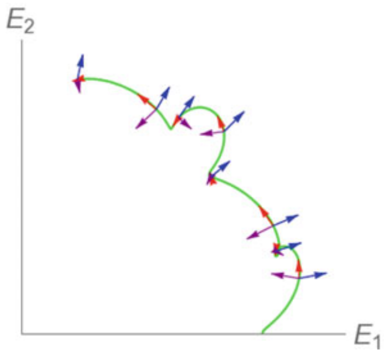
2. Show in $d = 3$ dimensions with $\omega(t) := i^{-1}(B(t)) \in \mathbb{R}^3$ (see (13.4.8)) that

$$C''(t) = O^{-1}(t)\, c''(t) - 2\omega(t) \times C'(t) - \omega(t) \times (\omega(t) \times C(t)) - \omega'(t) \times C(t).$$

3. Calculate the Coriolis force (see below) acting on a biker of mass $m = 100\,\text{kg}$ (including the bike!) that is moving westward in Berlin at a speed of $20\,\text{km/h}$. What is the amount of the horizontal component, and in which direction does it pull? ◊

If $c(t) \in \mathbb{R}^d$ is the position of a particle with mass $m > 0$ in $\mathbb{R}^d$, then by Newton's equation, the terms of (14.2.5), multiplied by m, can be viewed as forces which —from the point of view of the body coordinate system— act on the particle:

1. $-2mBC'$ (in 3D: $-2m\,\omega \times C'$) is called *Coriolis force*,
2. $-mB^2C$ (in 3D: $-m\,\omega \times (\omega \times C)$) is called *centrifugal force*, and
3. $-mB'C$ (in 3D: $-m\,\omega' \times C$) is called the *Euler force*.



Velocity (red), Coriolis force (purple, orthogonal to C), and centrifugal force (blue, radial) for the curve C from Figure 14.2.1

Substituting B into the Lagrange function $\mathcal{L}$ in the case of vanishing potential yields, in view of $A = O'(t) = O(t)B(t)O(t)^{-1}$, the result $\mathcal{L} = \frac{1}{2}\mathrm{tr}\big(B\tilde{I}B^\top\big)$.

14.3 Solution of the Equations of Motion

If the potential V is translation invariant, i.e., independent of its first argument $a \in \mathbb{R}^d$, then a does not enter into the Lagrange function (14.2.4).

Consequently, the center of mass of the rigid body moves along a straight line with constant velocity:

$$v(t) = v(0) \quad , \quad a(t) = a(0) + v(0)t \qquad (t \in \mathbb{R}),$$

and the motion within the center of mass system is determined[5] by the Lagrange function

$$\mathcal{L} : T\mathrm{SO}(d) \to \mathbb{R} \quad , \quad \mathcal{L}(O, A) = \tfrac{1}{2}\mathrm{tr}\big(\tilde{I} O^{-1} A^\top A O\big) - V(0, O)\,. \quad (14.3.1)$$

We will now study this situation in more detail.

14.9 Example (Heavy Top in 2 Dimensions)
For $d = 2$ dimensions, one has

$$\mathrm{SO}(2) = \left\{ O = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \,\middle|\, \varphi \in [0, 2\pi) \right\} \cong S^1 \quad \text{and} \quad T\mathrm{SO}(2) \cong S^1 \times \mathbb{R}\,,$$

because $\mathrm{Alt}(2, \mathbb{R}) = \left\{ A = \left(\begin{smallmatrix} 0 & -v \\ v & 0 \end{smallmatrix}\right) \mid v \in \mathbb{R} \right\} \cong \mathbb{R}$, see also Example A.44.
In this parametrization of the tangent space of SO(2), one has $A^\top A = v^2 \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, hence

$$\mathcal{L}(O, A) = \tfrac{1}{2}\mathrm{tr}\left(\tilde{I} O^{-1} A^\top A O\right) = \tfrac{1}{2}\mathrm{tr}\big(\tilde{I}\big)\, v^2 = \tfrac{1}{2} J\, v^2 \text{ with } J := \sum_{k=1}^{n} m_k \|q_k\|^2 > 0. \quad (14.3.2)$$

Thus the equation of motion is equivalent to the one for the Hamilton function

$$H : T^* S^1 \to \mathbb{R}\,,\; H(p, \varphi) = \frac{p^2}{2J} - W(\varphi) \quad \text{with } W(\varphi) := V\left(0, \left(\begin{smallmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{smallmatrix}\right)\right).$$

The corresponding Hamiltonian equation can be solved explicitly (see Theorem 11.11). Likewise, the case of a rigid body in $\mathbb{R}^3$ that rotates about a *stationary axis* can be reduced to this 2-dimensional problem.

If even $W = 0$, then the rigid body rotates about its center of mass with an angular velocity determined by the energy $E = H(p_0, \varphi_0)$, namely

$$\varphi(t) = \varphi_0 + \tfrac{p_0}{J} t = \varphi_0 \pm \tfrac{\sqrt{E}}{J} t\,. \qquad \diamondsuit$$

The important case in physics, where a rigid body in $d = 3$ dimensions rotates freely, seems to be more difficult already because its configuration space SO(3) cannot be parametrized by coordinates as easily as SO(2). The Euler angles are available as coordinates in this case.

However, we cannot hope that the expression for the kinetic energy simplifies just as in (14.3.1). This is because the matrix $A^\top A \in \mathrm{Sym}(3, \mathbb{R})$ will now not necessarily commute with the rotation matrix $O \in \mathrm{SO}(3)$. Instead we consider two special cases:

---

[5]We assume here and in the following deliberations on dynamics that the mass distribution is not degenerate.

- The *force free top*, which is the top in the potential $V = 0$.
- The *heavy symmetric top*: The tensor of inertia (14.2.3) of a *symmetric* top has two of its eigenvalues equal. We talk about a *heavy* top, if V describes a constant field of gravitation.

14.3.1 Force Free Top

In the case of a force free top, we transform into body coordinates.

In $d = 3$ dimensions, instead of the antisymmetric matrices $B \in \mathrm{Alt}(3, \mathbb{R})$ from Lemma 14.6, we use the more familiar vectors

$$\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \in \mathbb{R}^3 \quad \text{with} \quad i(\omega) = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ \omega_1 & -\omega_2 & 0 \end{pmatrix} = B \,,$$

see (13.4.8), to parametrize the tangent space of SO(3). Using the property $i(a)c = a \times c \quad (a, c \in \mathbb{R}^3)$ from (13.4.8) and the identity

$$\langle a \times c, b \times c \rangle = \langle a, b \rangle \, \|c\|^2 - \langle a, c \rangle \, \langle b, c \rangle \qquad \big(a, b, c \in \mathbb{R}^3\big)$$

for the cross product, with $a = b = \omega$ and $c = q_k$, the Lagrange functions is obtained in the form

$$\mathcal{L}(\omega) = \tfrac{1}{2}\mathrm{tr}\left(i(\omega)\tilde{I}i(\omega)^\top\right) = \tfrac{1}{2}\,\langle \omega, I\omega \rangle \,. \tag{14.3.3}$$

Here the symmetric matrix $I \in \mathrm{Sym}(3, \mathbb{R})$ is equal to $I = J1\!\!1 - \tilde{I}$ with $\tilde{I}$ from (14.2.3) and

$$J := \sum_{k=1}^{n} m_k \|q_k\|^2 \text{ , thus } I = \sum_{k=1}^{n} m_k \begin{pmatrix} q_{k,2}^2+q_{k,3}^2 & -q_{k,1}q_{k,2} & -q_{k,1}q_{k,3} \\ -q_{k,1}q_{k,2} & q_{k,1}^2+q_{k,3}^2 & -q_{k,2}q_{k,3} \\ -q_{k,1}q_{k,3} & -q_{k,2}q_{k,3} & q_{k,1}^2+q_{k,2}^2 \end{pmatrix}. \tag{14.3.4}$$

I is called the *tensor of inertia*. As the terms in the sum for I in (14.3.4) are positive semidefinite with kernel span(q_k) (and an eigenvalue $\|q_k\|^2$ of multiplicity two), I is itself positive semidefinite, and it is actually positive definite, except in the case when all mass points q_k are on a single line.

14.10 Exercise (Steiner's Theorem) Show that the tensor of inertia from (14.3.4), viewed as a function $I : \mathbb{R}^{3d} \to \mathrm{Sym}(3, \mathbb{R})$ of the n mass points $q = (q_1, \ldots, q_n)$, will be minimal[6] with respect to the translations

$$T_a(q) := (q_1 - a, \ldots, q_n - a) \qquad (a \in \mathbb{R}^3)$$

[6]For symmetric matrices, one has $A \leq B$ if $B - A$ is positive semidefinite.

if a is the center of mass of the q_N from (14.2.2). Also show that

$$I(q) \;=\; I\big(T_{q_N}(q)\big) \;+\; m_N\|q_N\|^2(1\!\!1_3 - P_N)\,,$$

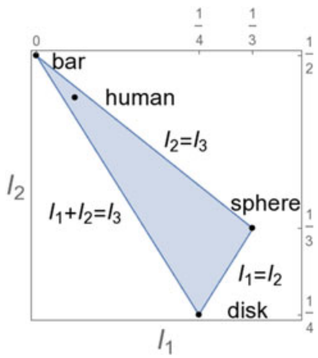
where m_N is the total mass and $q_N \in \mathbb{R}^3 \setminus \{0\}$, and P_N is the projection onto $\mathrm{span}(q_N)$. So the moment of inertia for rotation about the axis $\mathrm{span}(q_N)$ is the sum of the moments of inertia of a body of the same mass, shifted by its center of mass, and of a body concentrated in its center of mass. ◊

Usually the tensor of inertia I for a body is given with respect to its center of mass. The eigenvalues $0 \le I_1 \le I_2 \le I_3$ of I are also called *principal moments of inertia*, and one-dimensional eigenspaces for them are called *principal axes of inertia*. The latter are pairwise orthogonal, or else if some principal moments of inertia are equal, corresponding axes of inertia can at least be chosen to be pairwise orthogonal.

14.11 Exercise (Principal Moments of Inertia)
Show that for $n \ge 3$ mass points $q_k \in \mathbb{R}^3$, the cone

$$\big\{(I_1, I_2, I_3) \in \mathbb{R}^3 \mid I_1 \le I_2 \le I_3 \le I_1 + I_2\big\}$$

is the range for the principal moments of inertia. Which range is covered if $n = 3$, assuming the center of mass q_N is at the origin?
By passing from the sum to an integral in (14.3.4), confirm the labels for the corners in the figure of principal moments of inertia on the right.



In this figure we have assumed that $I_1 + I_2 + I_3 = 1$.
The data for a human refer to the person's center of mass for the average adult, as standing upright, and can vary considerably. ◊
The interesting feature of (14.3.3) is that only three of the six phase space coordinates, namely the entries of angular velocity ω, remain yet.

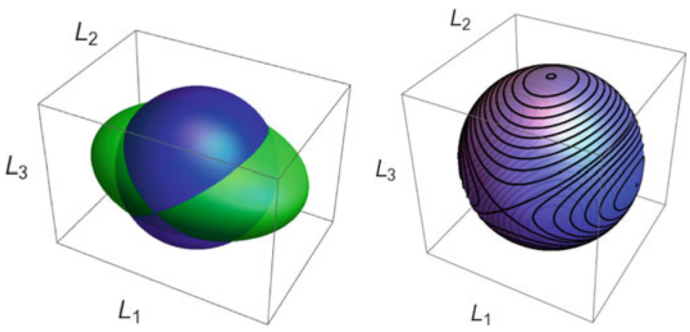


Figure 14.3.1 Left: Intersection of the ellipsoids $M^{-1}(m)$ (blue) and $\mathcal{L}^{-1}(e)$ (green). Right: Equipotential curves on $M^{-1}(m)$. The axes are the components of the angular momentum

14.12 Theorem

- *$\mathcal{L} : T\mathrm{SO}(3) \to \mathbb{R}$ from (14.3.3) is a constant of motion.*

- *The total angular momentum of the force free rigid body in spatially fixed coordinates, i.e., the restriction $L : T\mathrm{SO}(3) \to \mathbb{R}^3$ of the mapping*

$$\tilde{L} : \mathbb{R}^{3n}_q \times \mathbb{R}^{3n}_v \to \mathbb{R}^3 \quad , \quad \tilde{L}(q, v) = \sum_{k=1}^{n} m_k q_k \times v_k \, ,$$

is constant in time. In body coordinates, it is of the form $\mathbf{L} = \mathbf{I}\,\omega$, and $\|\mathbf{L}\|$ is a constant of motion for the force free top.

Proof:

- The Lagrange function $\mathcal{L}$ is a constant of motion since in the present case, due to the equation $H(p, q) = \langle p, v\rangle - \mathcal{L}(q, v)$ from Theorem 8.6, it is equal to the pull-back of the Hamilton function $H : T^*\mathrm{SO}(d) \to \mathbb{R}$ to $T\mathrm{SO}(d)$.[7]

- With (14.3.4), it follows that

$$\tilde{L}(q, v) = \textstyle\sum_{k=1}^{n} m_k q_k \times (\omega \times q_k) = \sum_{k=1}^{n} m_k \big(\|q_k\|^2 \omega - \langle \omega, q_k\rangle\, q_k\big) = I\omega \, .$$

- The angular momentum is constant because of Noether's Theorem (Theorem 13.22):

The diagonal action of the rotation group on the configuration space, i.e.,

$$\tilde{\Phi} : \mathrm{SO}(3) \times \mathbb{R}^{3n}_q \to \mathbb{R}^{3n}_q \quad , \quad \tilde{\Phi}_O(q_1, \ldots, q_n) = \big(Oq_1, \ldots, Oq_n\big) \, ,$$

can be restricted to the orbit, which is isomorphic to SO(3), and thus becomes a left action on SO(3). The Lagrange function is invariant under the lift of this left action to $T\mathrm{SO}(3)$, because

$$\big\langle O\omega,\ OIO^{-1}\ O\omega\big\rangle = \langle \omega,\ I\omega\rangle \qquad \big(O \in \mathrm{SO}(3)\big).$$

Analogously, the Hamilton function $H : T^*\mathrm{SO}(d) \to \mathbb{R}$ is invariant under the cotangent lift Φ of the left action. The angular momentum $L : T\mathrm{SO}(3) \to \mathbb{R}^3$, pulled back to $T^*\mathrm{SO}(3)$, is a momentum mapping of Φ (in analogy to Example 13.19), and the Hamilton function H is Φ-invariant.

- $\|\mathbf{L}\|$ is constant in time because $\mathbf{L}(t) = O^{-1}(t)L$. □

[7] While Theorem 8.6 was not worded for configuration manifolds M, but rather just for open subsets U of $\mathbb{R}^n$, it can still be read as a statement in the image of each chart of M.

Next to $\mathcal{L}$, the square $M := \langle L, L\rangle$ of the angular momentum vector is a positive definite quadratic form on $\mathbb{R}^3_\omega$. If the eigenvalues of I are pairwise distinct, then typical values of $\mathcal{L}$ and M define bounded curves in the space $\mathbb{R}^3_\omega$. So we have solved the differential equation for the angular velocity ω, except for the parametrization by time, see Figure 14.3.1.

We determine the evolution of the angular momentum $\mathbf{L}(t) = O(t)^{-1}L(t) \in \mathbb{R}^3$ in the moving basis. According to Lemma 14.6 and Theorem 14.12, one obtains the Euler equation

$$\dot{\mathbf{L}}(t) = \mathbf{L}(t) \times \Omega(t)$$

for the force free top. Since $\mathbf{L}(t) = \mathbf{I}\,\Omega(t)$, this is a differential equation for the vector Ω:

$$\mathbf{I}\dot{\Omega}(t) = \big(\mathbf{I}\,\Omega(t)\big) \times \Omega(t)\,.$$

If the symmetric matrix $\mathbf{I}$ is diagonal, $\mathbf{I} = \mathrm{diag}(\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3)$, then the coordinate form of the Euler equation for $\Omega = \begin{pmatrix}\Omega_1\\ \Omega_2\\ \Omega_3\end{pmatrix}$ is

$$\begin{aligned}
\mathbf{I}_1\,\dot{\Omega}_1 &= (\mathbf{I}_2 - \mathbf{I}_3)\,\Omega_2\,\Omega_3\,,\\
\mathbf{I}_2\,\dot{\Omega}_2 &= (\mathbf{I}_3 - \mathbf{I}_1)\,\Omega_3\,\Omega_1\,,\\
\mathbf{I}_3\,\dot{\Omega}_3 &= (\mathbf{I}_1 - \mathbf{I}_2)\,\Omega_1\,\Omega_2\,.
\end{aligned}$$

Equivalently, the coordinates $\begin{pmatrix}\mathbf{L}_1\\ \mathbf{L}_2\\ \mathbf{L}_3\end{pmatrix} := \mathbf{L}$ of the angular momentum in the moving basis satisfy $\mathbf{L}_k = \mathbf{I}_k\Omega_k$, hence

$$\begin{aligned}
\dot{\mathbf{L}}_1 &= \left(\tfrac{1}{\mathbf{I}_3} - \tfrac{1}{\mathbf{I}_2}\right)\mathbf{L}_2\,\mathbf{L}_3\,,\\
\dot{\mathbf{L}}_2 &= \left(\tfrac{1}{\mathbf{I}_1} - \tfrac{1}{\mathbf{I}_3}\right)\mathbf{L}_3\,\mathbf{L}_1\,,\\
\dot{\mathbf{L}}_3 &= \left(\tfrac{1}{\mathbf{I}_2} - \tfrac{1}{\mathbf{I}_1}\right)\mathbf{L}_1\,\mathbf{L}_2\,,
\end{aligned}$$

with the conserved quantities $E := \sum_{k=1}^{3}\frac{\mathbf{L}_k^2}{\mathbf{I}_k}$ and $\|\mathbf{L}\|^2 = \mathbf{L}_1^2 + \mathbf{L}_2^2 + \mathbf{L}_3^2$.

Figure 14.3.2 Left: A model of the Mir space station. Center: damaged space station. Right: the astronaut Michael Foale (Pictures courtesy of: The Mathematica Journal, Special Section: Dynamic Rotation of Space Station Mir, Oct 1999 (left), NASA/JPL/Space Science Institute (center) and The Mathematical Sciences Research Institute (MSRI, Berkeley, California), DVD The Right Spin (right))

The Right Spin

This is the title of a DVD published by the mathematician Robert Osserman about a serious accident of the *Mir* space station (Figure 14.3.2, left and center).

On May 25, 1997, while the astronauts were attempting to navigate the docking maneuver manually, without the automated navigation system, *Mir* collided with the unmanned supply module *Progress*. (After the disintegration of the Soviet Union, Ukraine charged for the use of this system, and the Russian space program was under fiscal austerity measures.)

To avoid the loss of pressure following the collision, the crew had to escape into one module of the space station. The next problem was a power outage, because the cosmonauts were forced to cut apart power cables in order to close a hatch of their module. The still remaining solar cells would no longer produce electricity because the collision had caused the station to rotate.

In the movie, the US astronaut Michael Foale (Figure 14.3.2 right), reports how he and his Russian colleagues attempted to stabilize *Mir*. To this end, one had to estimate (based on a broken model of *Mir*) which of the principal axes were stable.

The conjecture turned out to be wrong, and Foale attempted to solve the Euler equations for the station on a laptop whose batteries were beginning to run low. The attempts at a stabilization were successful enough for the crew to survive so that they could return to earth three and a half months later in a space shuttle.

14.13 Definition
The **moving ellipsoid of inertia** *of a rigid body with kinetic energy* $h > 0$ *is*

$$\mathbf{E}_h := \left\{\omega \in \mathbb{R}^3 \mid \tfrac{1}{2}\langle \omega, \mathbf{I}\,\omega\rangle = h\right\}.$$

It is on this ellipsoid that the vector $t \mapsto \Omega(t)$ in body coordinates has to move if the energy equals h. Analogously, the angular velocity Ω in spatially fixed coordinates satisfies

$$\Omega(t) \in \mathrm{E}_h(t) := O(t)\,\mathbf{E}_h \qquad (t \in \mathbb{R}).$$

14.14 Theorem (Poinsot) *The time dependent ellipsoid of inertia* $t \mapsto E_h(t)$ *of the spatially fixed coordinate system rolls without sliding on the planes that are orthogonal to the angular momentum vector* $\ell \in \mathbb{R}^3$*, i.e., on*

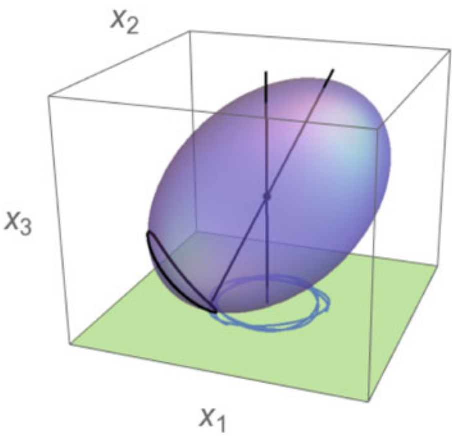
$$U_\pm := \{\omega \in \mathbb{R}^3 \mid \langle \omega, \ell\rangle = \pm 2h\}.$$

Proof:

- $E_h(t)$ meets the planes at $\pm\Omega(t)$, because $\ell = I(t)\Omega(t)$ implies $\langle \Omega(t), \ell\rangle = 2h$.
- $E_h(t)$ is tangential to these planes because the gradient of the kinetic energy at $\Omega(t) \in E_h(t)$ equals ℓ.
- The ellipsoid of inertia is not sliding as it moves on $U_\pm$, because the velocity of a point moving along the ellipsoid of inertia, $\omega(t) = O(t)\omega(0) \in E_h(t)$, is

$$\frac{\mathbf{d}O}{\mathbf{d}t}(t)\,\omega(0) = i\big(\Omega(t)\big)\,\omega(t)$$

(with the i from (13.4.8)). For $\omega(t) = \Omega(t)$, this expression vanishes. □



The ellipsoid of inertia rolls without sliding on the plane U_-

As we can see from this description, but also already from the intersection of the ellipsoids shown in Figure 14.3.1, motion about the axis for the middle moment of inertia is unstable.

14.15 Remark (Tennis Racket Theorem)
Among other conclusions, this leads to the following observation (see CUSHMAN and BATES [CB], Chapter III.8): If you hold a tennis racket horizontally and attempt to rotate it about its vertical axis, then the handle moves approximately horizontally. If you catch the racket again after it has performed one rotation, the top and bottom part will be exchanged, see Figure 14.3.3.b on page 382. ◊

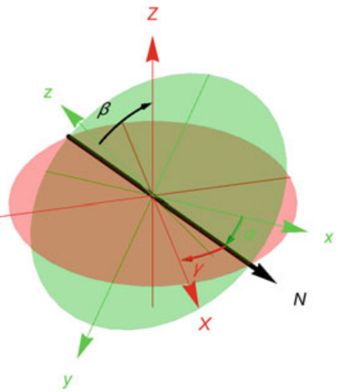
14.3.2 Heavy (Symmetric) Tops

As already in the case of the force free top, we will not assume that the top rotates about its center of mass already. To the contrary, a homogenous gravitational field will only show an effect if the point of suspension does *not* coincide with the center of mass.

The *Euler angles* are a parametrization of the rotation group SO(3). A rotation matrix $R \in \mathrm{SO}(3)$ is written as a product $D_3(\gamma)D_1(\beta)D_3(\alpha)$, with the rotations

$$D_1(\delta) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\delta) & \sin(\delta) \\ 0 & -\sin(\delta) & \cos(\delta) \end{pmatrix},$$

$$D_3(\delta) := \begin{pmatrix} \cos(\delta) & \sin(\delta) & 0 \\ -\sin(\delta) & \cos(\delta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$



about the x- and z-axes respectively. If we denote the line of intersection between the xy-plane and its image under R as *nodal line* N, then $\alpha \in [0, 2\pi)$ (resp. $\gamma \in [0, 2\pi)$) will be the angle between the x-axis (resp. its image, the X-axis) and N. $\beta \in [0, \pi)$ parametrizes the angle between both planes. α and β are therefore the geographic longitude and latitude of the Z-axis. Clearly the representation becomes degenerate if $\beta = 0$. See figure.

Applied to the heavy top, gravitation acts in the negative z-direction, so the xy-plane is horizontal. We assume that the XYZ-axes are rigidly attached to the body and are the principal axes of inertia for the moments of inertia I_1, I_2 and I_3. These I_i are to be calculated with respect to the point of suspension, which we take to be the origin.

The Lagrange function $L : T\mathrm{SO}(3) \to \mathbb{R}$ has $\mathcal{L}(\omega) = \frac{1}{2}\langle \omega, I\omega \rangle$ from (14.3.3) as its kinetic term. The potential energy in the homogenous gravitational field is the one of the center of mass. We assume that the point of suspension[8] lies on the axis through the center of mass, which points in Z-direction, in distance a. Then its z-component is $a\cos(\beta)$, and with acceleration of gravity $g > 0$, the potential energy is $V \equiv V(\beta) = ag\cos(\beta)$. We will use units of measurement for which $ag = 1$.

In terms of Euler angles (α, β, γ), the Lagrange function L reads

$$L\big(\alpha, \beta, \gamma, \dot{\alpha}, \dot{\beta}, \dot{\gamma}\big) = T\big(\alpha, \beta, \gamma, \dot{\alpha}, \dot{\beta}, \dot{\gamma}\big) - \cos(\beta) \qquad (14.3.5)$$

with kinetic energy

[8] which we have taken to be the origin of the cartesian body coordinate system.

$$T\big(\alpha,\beta,\gamma,\dot\alpha,\dot\beta,\dot\gamma\big)=\tfrac12\Big(I_1\big(\dot\alpha\sin(\beta)\sin(\gamma)+\dot\beta\cos(\gamma)\big)^2 +I_2\big(\dot\alpha\sin(\beta)\cos(\gamma)-\dot\beta\sin(\gamma)\big)^2+I_3\big(\dot\gamma+\dot\alpha\cos(\beta)\big)^2\Big)\,.$$

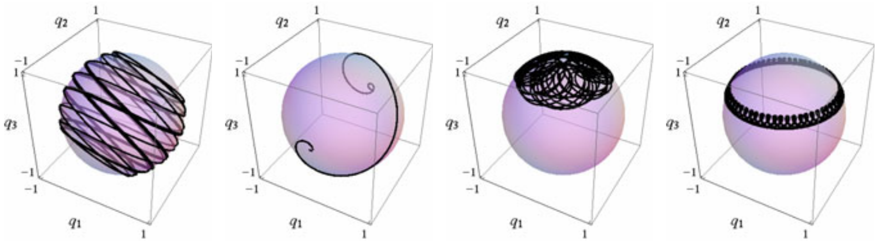


Figure 14.3.3 Motion of the locus of the I_3-axis on the unit sphere. From left to right: (a) no symmetry, no gravity; (b) tennis racket theorem; (c) no symmetry, with gravity; (d) symmetry, with gravity

From the numerical solution, it seems that the motion in the case of generic principal moments of inertia is not integrable[9] (see Figure 14.3.3c). In the axially symmetric case $I_1 = I_2$, the Lagrange function (14.3.5) takes a simpler form, namely[10]

$$L\big(\alpha,\beta,\gamma,\dot\alpha,\dot\beta,\dot\gamma\big)=\tfrac12\Big[I_1\big((\dot\alpha\,\sin(\beta))^2+\dot\beta^2\big)+I_3\big(\dot\gamma+\dot\alpha\cos(\beta)\big)^2\Big]-\cos(\beta)\,. \quad (14.3.6)$$

Here the coordinates α and γ do not appear any more. Then their conjugate momenta

$$p_\alpha=\frac{\partial L}{\partial\dot\alpha}=\dot\alpha\big(I_1\sin^2(\beta)+I_3\cos^2(\beta)\big)+\dot\gamma I_3\cos(\beta)\,,\ p_\gamma=\frac{\partial L}{\partial\dot\gamma}=\big(\dot\gamma+\dot\alpha\cos(\beta)\big)I_3 \quad (14.3.7)$$

are conserved quantities. The values ℓ_z and ℓ_Z of these conserved quantities correspond to the angular momenta about the vertical axis and about the fixed symmetry axis of the body. Together with the value E of the total energy $T+V$, we thus have three independent constants of motion, and the motion is integrable in the sense of Definition 13.2. From the relation $\ell_z-\ell_Z\cos(\beta)=\dot\alpha I_1\sin^2(\beta)$ we conclude the implicit first order differential equation $E=\frac12 I_1\dot\beta^2+V_{\mathrm{eff}}(\beta)$ with the effective potential

$$V_{\mathrm{eff}}(\beta):=\frac{\big(\ell_z-\ell_Z\cos(\beta)\big)^2}{2I_1\sin^2(\beta)}+\frac{\ell_Z^2}{2I_3}+\cos(\beta)\,.$$

Transforming to the variable $u:=\cos(\beta)$ with $\dot\beta=-\dot u/\sin(\beta)$ makes the equation more transparent. Because then,

$$\dot u^2=U_{\mathrm{eff}}(u)\quad\text{with}\quad U_{\mathrm{eff}}(u):=\frac{2EI_3-\ell_Z^2}{I_1I_3}(1-u^2)-\frac{(\ell_z-\ell_Z u)^2}{I_1^2}-\frac{2u(1-u^2)}{I_1},$$

hence the effective potential becomes a third degree polynomial.

---

[9]This can be proven, see MACIEJEWSKI and PRZYBYLSKA [MP], and references therein.

[10]This form does not assume that I_3 is the largest of the principal moments of inertia.

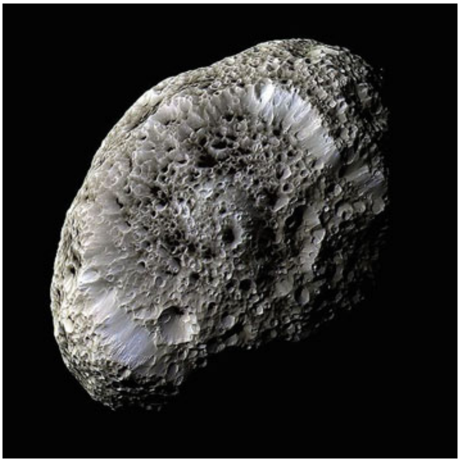
From $\lim_{u\to\pm\infty} U_{\text{eff}}(u) = \pm\infty$ and $U_{\text{eff}}(\pm 1) = -(\ell_z \mp \ell_Z)^2/I_1^2 \le 0$, we conclude that for $\ell_z \neq \pm\ell_Z$ and physically possible values E of the energy, the polynomial has two zeros in $(-1, 1)$. So the motion of the angle β occurs between two limit angles $0 < \beta_1 \le \beta_2 < \pi$, see Figure 14.3.3d. This motion of the top is called *nutation*.

The nutation is superimposed to the *precession* of α and the *rotation* of γ. Their differential equations are obtained by plugging the nutation $t \mapsto \beta(t)$ into (14.3.7).

14.16 Literature In [Whi], Chapter 6 by WHITTAKER (which was *the* monograph on analytical mechanics at the beginning of the 20th century), one can find the explicit solutions (in terms of elliptic functions).
Global aspects are discussed in [CB] by CUSHMAN and BATES. ◊

14.17 Exercise (Fast Top) Analyze at which frequency of rotation the upper equilibrium $\beta = 0$ of the heavy symmetric top becomes Lyapunov-stable. ◊

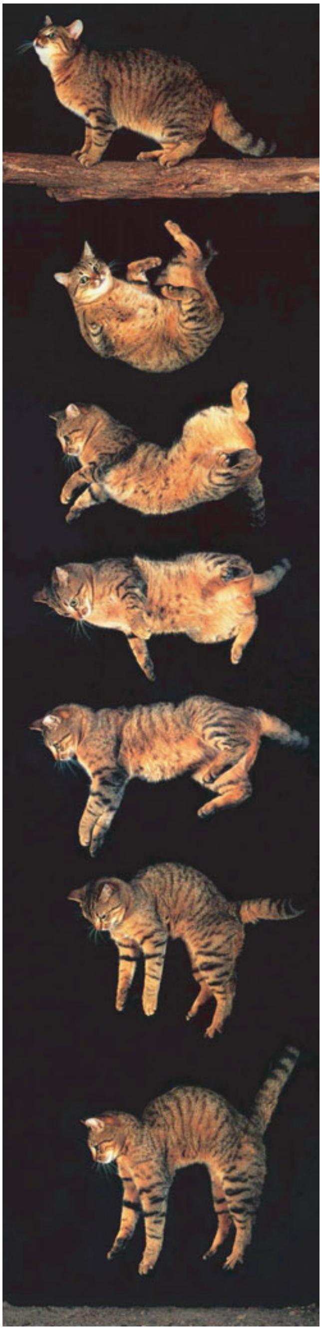
14.18 Remark (Hyperion)
Hyperion is a moon of the planet Saturn.The spacecraft Voyager 2 took its picture (see the figure[11]), and it was possible to determine the principal moments of inertia $I_1 < I_2 < I_3$. With a value of $(I_2 - I_1)/I_3 \approx 0.24$, they deviate significantly from the moments of inertia of an axially symmetric shape. Combined with a fairly eccentric orbit around Saturn ($\varepsilon \approx 0.12$), this leads to a rather chaotic rotation of Hyperion, see [WPM] by WISDOM, PEALE and MIGNARD. Hyperion is the only known moon in the solar system that has such a property.



In [ZP], ZUREK and PAZ estimated that after about 20 years, a quantum mechanical calculation of the rotation would significantly differ from the classical data. So Hyperion should be in a macroscopic quantum state. It was the content of the *Hyperion dispute* how it could be explained that we do not observe this macroscopic quantum state. ◊

[11] Picture: courtesy of NASA/JPL-Caltech.

14.4 Nonrigid Bodies, Nonholonomic Systems



The rigid bodies from Definition 14.3 consist of n particles in $\mathbb{R}^d$ with fixed distances. So their configuration space is a submanifold of $\mathbb{R}^{nd}$ whose dimension is no larger than the dimension of the group $\mathbb{SE}(d)$.
We will now generalize this framework of holonomic constraints in two directions:

1. As mentioned in Remark 8.10, holonomic constraints can be viewed as integrable *distributions*. More generally, one can talk about (*nonholonomic*) constraints when a not necessarily integrable distribution is given. Of particular importance are here the constraints that are linear in the velocity. A first example is the ball that rolls without sliding.
2. We will also study bodies that are flexible within themselves rather than rigid. Even if no distances of particles are fixed, we obtain a geometric structure that is not yet contained in a pure n-body problem. Namely, we will assume that certain distances can be *controlled*, like for instance the position of a joint or a hinge in biology or technology. Then the configuration space will become the total space of a bundle over this controlled base manifold.

14.4.1 Geometry of Flexible Bodies

Let us begin with item #2. This geometry was worked out in the 1980s and is described, e.g., in [MMR] by MARSDEN, MONTGOMERY and RATIU as well as in [Mon1] by MONTGOMERY.
The paradigm is the free fall of a cat as depicted on the right.
Even though the cat cannot change her angular momentum, she will nevertheless succeed to land on her legs, if she has the time to maneuver as shown in the picture.[12]

[12]Picture: Gérard Lacz.

The technique she uses is to use a geometrically defined holonomy. With this technique, cats can survive the fall from the upper floors of a highrise building. We humans can at least turn on an office chair without touching the ground by using a similar technique.

Generally, we are discussing the case of a free and proper action by a Lie group G on a (configuration) manifold Q. The quotient $B := Q/G$ is then a manifold of dimension $\dim(Q) - \dim(G)$ according to Theorem E.36 and can be viewed as the base manifold of a principal bundle

$$\pi : Q \to B \quad , \quad \pi(q) = [q] \equiv G \cdot q \tag{14.4.1}$$

with standard fiber G (see Remark F.5).

14.19 Example (Shape Space of a Flexible Body)
When deriving the kinematics of rigid bodies in Section 14.2, we started with the diagonal action of the Lie group $G = \mathbb{SE}(d)$ on $\mathbb{R}^{nd}$. This action is not free on $\mathbb{R}^{nd}$; however, it follows from Example 14.4.3 that for $n \geq d$ particles, the restriction of the diagonal action to the dense, G-invariant subset[13]

$$Q := \big\{q = (q_1, \dots, q_n) \in \mathbb{R}^{nd} \mid \dim_{\mathrm{aff}}\big(\{q_1, \dots, q_n\}\big) \geq d - 1\big\}$$

of *configurations* is free. Since the condition defining Q is open, $Q \subseteq \mathbb{R}^{nd}$ is an nd-dimensional submanifold. Then the quotient $B = Q/\mathbb{SE}(d)$ is called the *shape space* of the *flexible body*. The points $q \in \pi^{-1}(b)$ in the fiber over $b \in B$ are the *configurations* of the *shape* b.

We can obtain local coordinates near $b = \pi(q) \in B$ in the following way: Assume that for $a_k := q_k - q_1$ ($k = 2, \dots, n$) already the vectors $a_2, \dots, a_d \in \mathbb{R}^d$ are linearly independent (this assumption is always satisfied if we allow ourselves to permute the indices $1, \dots, n$). We consider the matrix

$$A := (a_2, \dots, a_d) \in \mathrm{Mat}\big(d \times (d-1), \mathbb{R}\big)\,.$$

As a consequence of Gram-Schmidt orthogonalization (or QR decomposition), there is a unique matrix $O \equiv O(q) \in \mathrm{SO}(d)$ such that

$$U := OA \in \mathrm{Mat}\big(d \times (d-1), \mathbb{R}\big)$$

is upper trapezoidal (that is, $(U)_{i,k} = 0$ for $1 \leq k < i \leq d$) and has entries $(U)_{k,k} > 0$. By construction, U is independent of the action of $\mathbb{SE}(d)$, so we can

[13] Here $\dim_{\mathrm{aff}}\big(\{q_1, \dots, q_n\}\big)$ is the *affine* dimension of the subset $\{q_1, \dots, q_n\}$ of $\mathbb{R}^d$, that is, the dimension of the subspace spanned by the set of their differences $q_i - q_k \in \mathbb{R}^d$.

write $U \equiv U(b)$. Locally, $U(b)$ depends smoothly on b. Finally, the entries $(V)_{i,k}$, $1 \le i < k \le n$) of the upper trapezoidal matrix

$$V := (v_2, \dots, v_n) \in \mathrm{Mat}\big(d \times (n-1), \mathbb{R}\big) \quad \text{with} \quad v_k := O a_k$$

are local coordinates on the shape space B near b. ◊

14.20 Exercises (Euclidean Symmetries)

1. Show that the actions by $\mathbb{SE}(d)$ on $\mathbb{R}^{nd}$ and Q are proper (Definition E.35).
2. Show that the form sphere of the three-body problem as defined on page 274 can be viewed as the quotient of the configuration space of three particles by a group action.
   Hint: Combine the dilations $\mathbb{R}^2 \to \mathbb{R}^2$, $x \mapsto \lambda x$ $\big(\lambda \in (0, \infty)\big)$ with $\mathbb{SE}(2)$, and let the group thus obtained act on the configuration space (11.3.2). ◊

In Example 14.19, B parametrizes the shapes of a *very* flexible 'body'. In many applications, the possible shapes will be limited by the choice of a G-invariant submanifold of Q. In any case, we presume that the shape of the body can be controlled as a function of time by prescribing a curve $c : [0, 1] \to B$.

If we prescribe a connection on the principal bundle $\pi : Q \to B$, then a closed curve c corresponds to a holonomy, i.e., a motion of the body described by an element of the group G. For example, c could describe a swimming stroke. Then the holonomy would be the translation effected by performing the stroke.

If the total space is a Riemannian manifold (Q, g),[14] then such a connection can be defined as follows. The horizontal subspace at $q \in Q$ that complements the (vertical) fiber direction $V_q := \ker(D\pi_q) \subset T_q Q$ will be chosen as

$$H_q := \{h \in T_q Q \mid g_q(h, v) = 0 \text{ for all } v \in V_q\} .$$

If G acts by isometries, as it does in Example 14.19, then this defines a G-invariant connection (in the sense of Definition F.17).

These geometric structures correspond to the Hamiltonian mechanics of a *natural mechanical system*, which is a function on the phase space $P := TQ$ that is of the form

$$H(q, v) = \tfrac{1}{2} g_q(v, v) + V(q) .$$

This mechanical system defines, by means of the canonical symplectic form ω_0, pulled back with the musical isomorphism $\flat : TQ \to T^*Q$ (see page 503), a Hamiltonian vector field on P.

By Theorem 13.16, we obtain a symplectic action of the group G on P, linear on each fiber. If G not only acts isometrically on Q, but also leaves V invariant, this action is a symmetry of the dynamics. The constants of motion are given by the

[14] In Example 14.19, $g(v, v) = \sum_{k=1}^{n} \frac{m_k}{2} \langle v_k, v_k \rangle$ with the Euclidean metric $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^d$ and the mass $m_k > 0$ of the kth particle.

momentum mapping $J : P \to \mathfrak{g}^*$ (Definition 13.18). This latter can be written, in terms of the vector field $X_\xi : Q \to TQ$ of $\xi \in \mathfrak{g}$, as

$$J(q, v) = X_q^\top(v) \qquad \big((q, v) \in TQ\big)$$

(where the transpose $X_q^\top : T_q Q \to \mathfrak{g}^*$ of $\xi \mapsto X_\xi$ is defined by means of g, see MONTGOMERY [Mon1], Chapter 14.1). So the action is linear on each fiber.

Now we can *almost* interpret $J : TQ \to \mathfrak{g}^*$ as a connection form $A : TQ \to \mathfrak{g}$ on the bundle $\pi : Q \to B$. Only 'almost', because the values of J lie in the dual Lie algebra $\mathfrak{g}^*$ instead of $\mathfrak{g}$.

But this deficit can be fixed if we define the *moment of inertia* at $q \in Q$ as the symmetric bilinear form $\mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$, $(\xi, \eta) \mapsto g_q\big(X_\xi(q), X_\eta(q)\big)$, or equivalently as a linear mapping

$$\mathbb{I}(q) : \mathfrak{g} \to \mathfrak{g}^* \quad , \quad \xi \mapsto g_q\big(X_\xi(q), \cdot\big) .$$

14.21 Theorem *The Lie algebra valued 1-form*

$$A : TQ \to \mathfrak{g} \quad , \quad A(q) := \mathbb{I}(q)^{-1} J\big(q, \cdot\big)$$

is the connection form of the Ehresmann connection H on the principal bundle $\pi : Q \to B$, i.e., $\ker\big(A(q)\big) = H_q$. *In particular, $J(q, v)$ vanishes for horizontal directions v.*

Proof: See MONTGOMERY [Mon1], *Proposition 14.1.* □

Thus mechanical systems with vanishing J move under a controlled dynamics in such a way as is defined by the Ehresmann connection H.

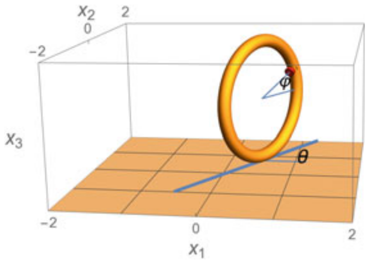
In the physics literature, the associated holonomy is often called *Berry phase* (in quantum mechanical phenomena) or *Hannay angle*. It can be shown that these quantities can be read off from the motion in many controlled systems in the limit of slowly changing parameters.

There is quite a variety of applications: The three-axes stabilization introduced in Example F.29 can be explained in this framework; likewise, the Foucault pendulum (depicted on page 241) and the cyclic locomotion of microorganisms; even a technique for swimming in a curved universe has been found (but it is not effective enough to escape a black hole, see AVRON and KENNETH [AvK]).

14.4.2 Nonholonomic Constraints

Tires, balls, etc. roll on a surface *without sliding* if their surface does not move at the point where it is in contact with the supporting surface. This leads to nonholonomic constraints that are linear in the velocities. For instance, we cannot drive sideways with a car or a bike, but we can still parallel park.

14.22 Example (Rolling Wheel) We consider a wheel that is rolling upright (see the article [BKMM] by Bloch, Krishnaprasad, Marsden and Murray). If the radius of the wheel is $r > 0$, then the hub is at point (x_1, x_2, r) just above the point of contact $(x_1, x_2, 0) \in \mathbb{R}^3$. Its present orientation in space is described by an angle θ. A second angle φ measures the position of a marked point on the wheel (say the location of the valve), relative to the point of contact with the ground. Thus the hub moves with a velocity vector $\dot{x} = \big(r\cos(\theta)\,\dot{\varphi}\,,\ r\sin(\theta)\,\dot{\varphi}\big)$ parallel to the surface.[15]



So the distribution $\mathcal{D} \subset TQ$ in the space $Q := \mathbb{R}^2_x \times S^1_\theta \times S^1_\varphi$ of positions, which is defined by the rolling process, is given by

$$\mathcal{D}(x, \theta, \varphi) = \big\{(\dot{x}, \dot{\theta}, \dot{\varphi}) \in \mathbb{R}^4 \mid \dot{x} = \big(r\cos(\theta)\dot{\varphi},\ r\sin(\theta)\dot{\varphi}\big)\big\} \qquad (14.4.2)$$

and it is obviously linear in the velocities. It is 2-dimensional and smooth. Abbreviating $q := (x_1, x_2, \theta, \varphi)$, it is spanned by the vector fields

$$X_i : Q \to TQ \quad , \quad X_1(q) := \big(r\cos(\theta), r\sin(\theta), 0, 1\big) \text{ and } X_2(q) := (0, 0, 1, 0)\ .$$

Thus their commutator $[X_1, X_2](q) = \big(r\sin(\theta), -r\cos(\theta), 0, 0\big)$, does not lie in $\mathcal{D}(q)$. So by the Theorem F.25 of Frobenius, it cannot be integrable.

Because $[X_2, [X_1, X_2]](q) = \big(r\cos(\theta), r\sin(\theta), 0, 0\big) = X_1(q) - (0, 0, 0, 1)$, the X_i and these two commutators span all of TQ already.
Therefore, by Chow's Theorem,[16] any two points in the configuration space Q can be connected by a smooth path $c : [0, 1] \to Q$ that is horizontal, i.e., whose velocity vector $\dot{c}(t)$ lies in $\mathcal{D}\big(c(t)\big)$. For example, it is possible to move the wheel from point A to point B in the plane in such a manner that its valve will point in the same direction at A and at B. ◊

Such kinematic considerations are complemented by considerations of the nonholonomic dynamics. To this end, a Lagrange function $L : TQ \to \mathbb{R}$ is restricted to the distribution $\mathcal{D} \subset TQ$, i.e., we set $L_\mathcal{D} := L\restriction_\mathcal{D}$. If $n := \dim(Q)$ and the distribution has rank $n - p$, then it can locally be written as a zero set of p independent 1-forms $\omega_1, \dots, \omega_p$. In local coordinates $q = (r, s)$ on Q with $s = (s_1, \dots, s_p)$, $r = (r_1, \dots, r_{n-p})$, one has

[15] We now suppress the vertical x_3 component.

[16] **Theorem (Chow):** If a distribution on a connected manifold Q is spanned by vector fields $X_1, \dots, X_n$ whose iterated commutators (including the vector fields themselves) in turn span all of TQ, then for all $q_0, q_1 \in Q$ there is a horizontal path $c \in C^1\big([0, 1], Q\big)$ with $c(i) = q_i$. See Montgomery, [Mon1], Chapter 2.

$$\omega_a(r,s) = ds_a + \sum_{\alpha=1}^{n-p} A_{a,\alpha}(r,s)\, dr_\alpha \qquad (a=1,\ldots,p). \tag{14.4.3}$$

Plugging this in and summing over α yields the restricted Lagrange function in coordinates:

$$L_{\mathcal{D}}(r,s,\dot r) = L\big(r,s,\dot r, -A_{\cdot,\alpha}(r,s)\,\dot r_\alpha\big)\,.$$

The *Lagrange-d'Alembert equations of motion* are obtained by variation of curves $t \mapsto q(t)$ that are tangential to the distribution, where the variation $\delta q(t)$ itself also lies in $\mathcal{D}\big(q(t)\big)$. In local coordinates, these equations are of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_{\mathcal{D}}}{\partial \dot r_\alpha} - \frac{\partial L_{\mathcal{D}}}{\partial r_\alpha} + A_{a,\alpha}\frac{\partial L_{\mathcal{D}}}{\partial s_a} = -\frac{\partial L}{\partial \dot s_a} B_{a,\alpha,\beta}\,\dot r_\beta \qquad (\alpha = 1,\ldots,n-p),$$

where B is the curvature of the connection defined by A. It is thus often called magnetic term (see Remark F.31.2.). In coordinates,

$$B_{\cdot,\alpha,\beta} = \frac{\partial A_{\cdot,\alpha}}{\partial r_\beta} - \frac{\partial A_{\cdot,\beta}}{\partial r_\alpha} + A_{a,\alpha}\frac{\partial A_{\cdot,\beta}}{\partial s_a} - A_{a,\beta}\frac{\partial A_{\cdot,\alpha}}{\partial s_a}\,.$$

14.23 Example (Rolling Wheel)
Assume that the wheel from Example 14.22 has an axially symmetric mass distribution with total mass $m > 0$. One principal axis of its tensor of inertia (14.2.3) will then coincide with the axle of the wheel, and $\tilde I$ is symmetric with respect to rotations orthogonal to the axle. If we denote the corresponding principal moment of inertia by I, and the other two by J, the Lagrange function will be

$$L : TQ \to \mathbb{R} \quad , \quad L\big(x,\theta,\varphi,\dot x,\dot\theta,\dot\varphi\big) = \tfrac{1}{2}\big(m\|\dot x\|^2 + I\dot\varphi^2 + J\dot\theta^2\big)\,.$$

The form of the distribution (14.4.2) suggests to use the angles as the r-coordinates in (14.4.3). The restricted Lagrange function is then

$$L_{\mathcal{D}}\big(x,\theta,\varphi,\dot x,\dot\theta,\dot\varphi\big) = \tfrac{1}{2}\big((mr^2+I)\dot\varphi^2 + J\dot\theta^2\big)\,.$$

Using the variable names as indices of A, one has $A_{x_1,\theta} = A_{x_2,\theta} = 0$ and $A_{x_1,\varphi}(x,\theta,\varphi) = -r\cos(\theta)$, $A_{x_2,\varphi}(x,\theta,\varphi) = -r\sin(\theta)$. It follows that

$$B_{x_1,\theta,\varphi}(x,\theta,\varphi) = r\sin(\theta) \quad , \quad B_{x_2,\theta,\varphi}(x,\theta,\varphi) = -r\cos(\theta)\,.$$

The equations of motion are therefore

$$(mr^2+I)\ddot\varphi = 0\,,\ J\ddot\theta = 0$$

with the solution

$$\varphi(t) = \varphi(0) + \dot{\varphi}(0)t\ , \theta(t) = \theta(0) + \dot{\theta}(0)t \qquad (t \in \mathbb{R}).$$

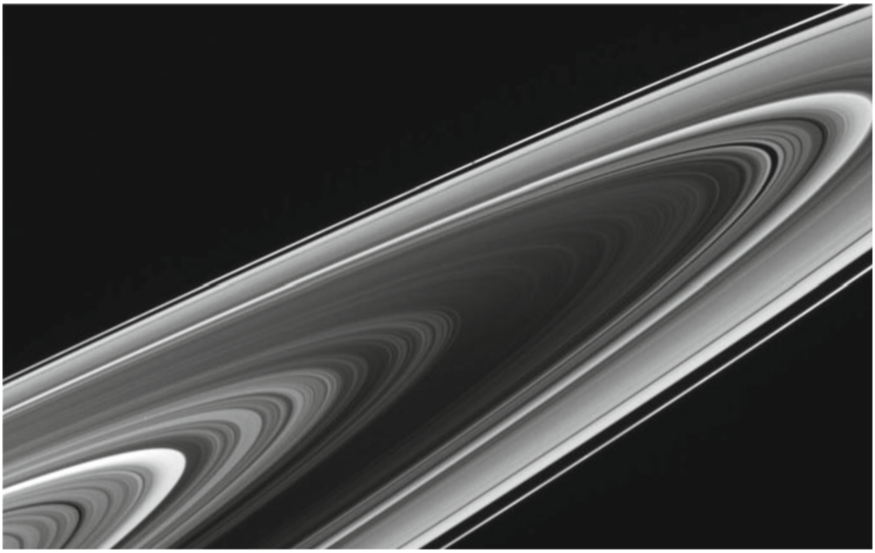
So, depending on the initial condition $\dot{\theta}(0)$, the wheel traverses uniformly either a circle or a straight line in the plane. ◊

While the motion of the wheel could have been guessed and does not fall outside the framework of Hamiltonian mechanics, unusual phenomena do occur already in nonholonomic systems that are not complicated:

- The phase space volume need no longer be conserved, even though there is no friction. This phenomenon can lead to asymptotic stability. Such asymptotic stability is observed in shopping carts that, when running unattended, orient themselves in such a way that the handle points to the front.
- In contrast to the symplectic case, there is a kind of Poisson bracket that does not always satisfy the Jacobi identity.
- In contrast to Noether's theorem, continuous symmetries need not lead to conserved quantities.

14.24 Literature In [BMZ], Bloch, Marsden and Zenkov present such cases and give an overview over the literature. ◊

Chapter 15
Perturbation Theory



The rings of Saturn, picture taken by the spacecraft Cassini in 2005.
Picture: courtesy of NASA/JPL-Caltech.

In perturbation theory, one considers dynamical systems whose solution is not known explicitly, but can be controlled by comparison with a known solution of a different dynamical system on the same phase space. In the Hamiltonian case, this approximation is particularly precise. In the extreme case of very irrational frequencies, the approximation is valid for all times.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_15

15.1 Conditionally Periodic Motion of a Torus

“Je n’avais pas besoin de cette hypothèse-là.” (Pierre-Simon Laplace)[1]

Not all Hamiltonian systems are integrable. For $n \geq 2$ degrees of freedom, non-integrability is even typical in a precise sense (namely *generic* in the sense of Remark 2.44.2), see MARKUS and MEYER [MaMe].

In nonintegrable systems, the motion within the $(2n-1)$-dimensional energy shell is not confined to n-dimensional tori for a positive Liouville measure of initial conditions, and it can look quite complex (‘chaotic’). Entirely new notions had to be developed to describe the long term behavior of the orbits, like for instance the notions from ergodic theory described in Chapter 9.

We now focus instead on systems that are *almost* integrable, which means the Hamiltonian is of the form

$$H_\varepsilon(I,\varphi) = H_0(I) + \varepsilon H_1(I,\varphi),$$

with $|\varepsilon|$ small. First the goal will be to describe the orbits for times that are not too large.

So we are looking at systems of differential equations of the form

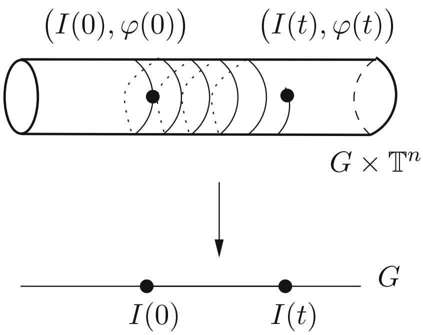
$$\begin{aligned} \dot I_k &= \quad 0 \quad + \varepsilon f_k(I,\varphi) \qquad \big(k \in \{1,\ldots,m\}\big) \\ \dot\varphi_l &= \omega_l(I) + \varepsilon g_l(I,\varphi) \qquad \big(l \in \{1,\ldots,n\}\big) \end{aligned} \tag{15.1.1}$$

on a phase space of the form $G \times \mathbb{T}^n$, with $G \subseteq \mathbb{R}^m$ open. Generalizing, we will not assume that the system is Hamiltonian. Also the number of action variables does not have to coincide with the number of angle variables.

In this context, it is better to talk about **slow** variables (I) and **fast** variables (φ), because obviously for every ε-independent time $T > 0$, one has:

$$\|I(t) - I(0)\| = \mathcal{O}(\varepsilon) \quad \text{and} \quad \|\varphi(t) - \varphi(0)\| = \mathcal{O}(1) \qquad \big(|t| \leq T\big).$$

It is sensible to view the phase space as a bundle $G \times \mathbb{T}^n$ over the base space G with the torus $\mathbb{T}^n$ as the fiber. In G, one has then only the slow drift of the I variables, see the figure to the right. In many cases, we are mainly interested in the time evolution of the slow variables I.
To investigate it, we will apply notions known from ergodic theory (Chapter 9.4) to equation (15.1.1):



Fast and slow variables

[1] “I did not need this hypothesis”. As a reply to Napoleon’s question why there was no mention of God in his book *Mécanique Céleste* (Celestial Mechanics), which among other things laid part of the foundation for perturbation theory. According to W. Rouse Ball: A short account of the history of mathematics, 4th edition (1908), page 418.

15.1 Definition

- *The* **space average** *of an integrable function* $h : G \times \mathbb{T}^n \to \mathbb{R}$ *is the function*

$$\langle h \rangle : G \to \mathbb{R} \quad , \quad \langle h \rangle(I) = \int_{\mathbb{T}^n} h(I, \varphi) \frac{\mathrm{d}\varphi}{(2\pi)^n} .$$

- *The* **time average** *of* h *with respect to* $\omega : G \to \mathbb{R}^n$ *is the function*

$$h^* : G \times \mathbb{T}^n \to \mathbb{R} \quad , \quad h^*(I, \varphi) = \lim_{T \to \infty} \frac{1}{T} \int_0^T h\big(I, \varphi + \omega(I)t\big) \, \mathrm{d}t .$$

We will denote the solution to (15.1.1) for given initial conditions $x_0 = (I_0, \varphi_0)$ as $t \mapsto I_\varepsilon(t, x_0)$. The *averaging principle* consists of comparing this solution with the solution $t \mapsto J_\varepsilon(t, x_0)$ of a simpler differential equation, called the averaged system

$$\dot{J}_k = \varepsilon \, \langle f_k \rangle (J) \qquad (k = 1, \ldots, m)$$

with initial value $J_\varepsilon(0, x_0) := I_0$. The ε-dependence of the averaged system is simple: $J_\varepsilon(t, x_0) = J_1(\varepsilon t, x_0)$.

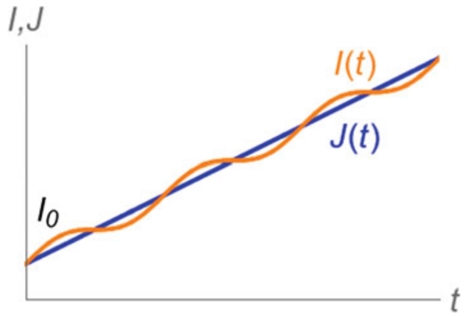
Under certain hypotheses, the difference $\| I_\varepsilon(t, x_0) - J_\varepsilon(t, x_0) \|$ will be small for a long time interval.

15.2 Example (Averaging Principle)
For $f : \mathbb{R} \times S^1 \to \mathbb{R}$, $f(I, \varphi) := 1 + \cos(\varphi)$ and $\omega \in \mathbb{R} \backslash \{0\}$, let

$$\dot{I} = \varepsilon f(I, \varphi) \quad , \quad \dot{\varphi} = \omega$$

(hence $g \equiv 0$ in (15.1.1)). The solution of the initial value problem is

$$I_\varepsilon(t, x_0) = I_0 + \varepsilon t + \frac{\varepsilon}{\omega} \sin(\omega t + \varphi_0)$$
$$\varphi_\varepsilon(t, x_0) = \varphi_0 + \omega t .$$



The averaged 'system' (in reality we only have one variable left) is

$$\dot{J} = \varepsilon \langle f \rangle (J) = \varepsilon \quad \text{, thus} \quad J_\varepsilon(t, x_0) = I_0 + \varepsilon t .$$

In this case, one has

$$|I_\varepsilon(t, x_0) - J_\varepsilon(t, x_0)| \le \frac{\varepsilon}{\omega} = \mathcal{O}(\varepsilon) ,$$

uniformly in the times $t \in \mathbb{R}$ and the initial values x_0, see figure on page 393. However, the distance diverges as $\omega \to 0$, and this is the typical behavior. ♢

Of course, in this simple example, we can solve the system by quadratures and do not need to rely on an approximate solution. Nevertheless, this example shows us what is essential. While the slow variable I oscillates 'rapidly' with a fluctuation of order $\mathcal{O}(\varepsilon)$, due to the φ-dependence of f, it is only the space average $\langle f \rangle = 1$ of f that contributes to a long term change in f. Here this space average corresponds exactly to the time average

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f\big(I_\varepsilon(t,x_0),\varphi_\varepsilon(t,x_0)\big)\,\mathrm{d}t = \lim_{T\to\infty}\frac{1}{T}\int_0^T \big(1+\cos(\varphi_0+\omega t)\big)\,\mathrm{d}t$$
$$= \lim_{T\to\infty}\frac{1}{T}\left(T + \frac{\sin(\varphi_0+\omega T)-\sin(\varphi_0)}{\omega}\right) = 1\,.$$

If the space and time averages of f were to coincide in the general case, we could be sure that the averaging principle gives us a good approximation to the evolution of the slow variables (assuming a fast convergence of the time average). However, this is unfortunately not always the case.

15.3 Example (Failure of the Averaging principle)
On the phase space $\mathbb{R}^2 \times \mathbb{T}^2$, we consider the differential equation

$$\dot I_1 = -\varepsilon\,\sin(\varphi_1 - 2\varphi_2),\ \dot I_2 = \varepsilon\,\big(\cos(\varphi_1-2\varphi_2)+\sin\varphi_2\big),\ \dot\varphi_1 = 2,\ \dot\varphi_2 = 1 + I_1.$$

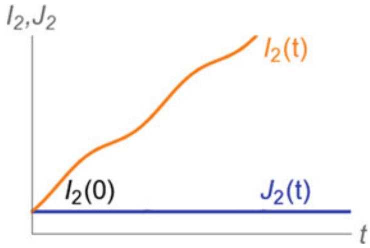
So the averaged system is

$$\dot J_1 = \dot J_2 = 0\,.$$

For the initial values with

$$\varphi_1(0) = \varphi_2(0) = 0 \text{ and } I_1(0) = 0\,,$$

we denote the solution simply as (I,φ):



$$\varphi_1(t) = 2t \quad,\quad \varphi_2(t) = t$$
$$I_1(t) = 0 \quad,\quad I_2(t) = I_2(0) + \varepsilon\,\big(t - \cos(t) + 1\big).$$

In contrast to the slow variables, the averaged variables remain constant:

$$J_1(t) = I_1(0) = 0 \quad\text{and}\quad J_2(t) = I_2(0)\,.$$

So in the limit of large times $|t|$, one has $\|I(t) - J(t)\| = \varepsilon|t - \cos(t) + 1| \sim \varepsilon|t|$, see the figure.

Thus, applying the averaging principle does not lead to a good approximation of the actions I_k by the averaged actions J_k in this case, at least not for the given initial conditions. ◊

If we look for conditions under which we can successfully apply the averaging principle, we need to analyze why, in the second example, the difference between I_2 and J_2 is large of order 1 already for times $t = \mathcal{O}(1/\varepsilon)$. To this end, we study the motion

$$\varphi_1(t) = 2t \quad , \quad \varphi_2(t) = t$$

on the torus $\mathbb{T}^2$, which occurs in Example 15.3. This motion is 2π-periodic, see Figure 15.1.1.

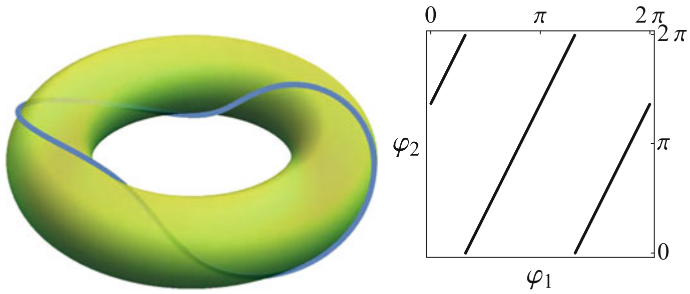


Figure 15.1.1 2π-periodic motion on $\mathbb{T}^2$

So we cannot expect for the time average and the space average of a function f on $\mathbb{T}^2$ to be equal.

In particular, the coefficient $f := f_2$ of the system of differential equations (with $f_2(I, \varphi) = \cos(\varphi_1 - 2\varphi_2) + \sin(\varphi_2)$) satisfies

$$\langle f_2 \rangle (I) \equiv 0\,,$$

but

$$\lim_{t\to\infty} \frac{1}{T} \int_0^T f_2\big(I, \varphi(t)\big)\,\mathrm{d}t = 1\,.$$

So we first need to look for conditions that enforce the equality of space and time averages for a motion on the torus.

15.4 Definition

- *Let $\varphi = (\varphi_1, \ldots, \varphi_n)$ be angle coordinates on the torus $\mathbb{T}^n$. Then for $\omega = (\omega_1, \ldots, \omega_n) \in \mathbb{R}^n$, the flow $\Phi : \mathbb{R} \times \mathbb{T}^n \to \mathbb{T}^n$ generated by the translation invariant vector field $\varphi \mapsto \omega$ is called* **conditionally periodic motion**.
- *The $\omega_l \in \mathbb{R}$ are called the* **frequencies** *of the conditionally periodic motion.*

- *They are called (rationally)* **independent** *if, for* $k \in \mathbb{Z}^n$*, the equation* $\langle k, \omega\rangle = 0$ *implies* $k = 0$.

15.5 Exercise (Conditionally Periodic Motion)
How often do the hour hand and the minute hand meet during one day? ◊

The notion of rational independence has already played a role in the harmonic oscillator, see Definition 6.31. Obviously, the flow on $\mathbb{T}^n \pmod{2\pi}$ is $\pmod{2\pi}$

$$\Phi_t\big(\varphi(0)\big) \equiv \varphi(t) = \big(\varphi_1(t), \ldots, \varphi_n(t)\big) = \big(\varphi_1(0) + \omega_1 t, \ldots, \varphi_n(0) + \omega_n t\big)\,. \tag{15.1.2}$$

In Theorem 5 of [Wey], Hermann Weyl stated as much as the following in 1916:

15.6 Theorem (Weyl)

- *For continuous functions* $f : \mathbb{T}^n \to \mathbb{C}$*, space and time averages of the conditionally periodic flow exist.*
- *If the frequencies* ω_l *are rationally independent, then space and time averages of (not necessarily continuous) Riemann integrable*[2] *functions are equal, i.e.,*

$$f^*(\varphi) = \langle f\rangle \qquad (\varphi \in \mathbb{T}^n).$$

The proof of this theorem will be given at the end of this chapter.

15.7 Remark (Independence)
If $n = 1$ as in our first Example 15.2, then independence of ω is equivalent to $\omega \neq 0$. This condition is satisfied in Example 15.2.

If however $n = 2$ as in Example (15.3), then independence of ω_1 and ω_2 is equivalent to $\omega_2 \neq 0$ and $\omega_1/\omega_2 \notin \mathbb{Q}$. This condition is violated in Example 15.3 for $I_1(0) = 0$, because then $\omega_1/\omega_2 = 2$. ◊

15.8 Corollary (Independence) *If the frequencies are independent, then*

- *every solution* $t \mapsto \varphi(t, \varphi_0)$ *on* $\mathbb{T}^n$ *is* **equidistributed**. *This means:*
  For every Jordan measurable[3] *set* $U \subseteq \mathbb{T}^n$*, the average time which the solution stays in* U*, namely* $\lim_{T\to\infty} T^{-1}\int_0^T \mathbb{1}_U\big(\varphi(t, \varphi_0)\big)\,\mathrm{d}t$ *equals the Haar measure of* U.
- *In particular, every orbit is dense on* $\mathbb{T}^n$.

Proof:

- By Theorem 15.6, the Riemann-integrable characteristic function $\mathbb{1}_U$ satisfies

$$\lim_{T\to\infty} T^{-1}\int_0^T \mathbb{1}_U\big(\varphi(t, \varphi_0)\big)\,\mathrm{d}t = \mathbb{1}_U^*(\varphi_0) = \langle \mathbb{1}_U\rangle = \frac{\lambda^n(U)}{\lambda^n(\mathbb{T}^n)} = \frac{\lambda^n(U)}{(2\pi)^n}\,.$$

---

[2] A corresponding statement is not true for every *Lebesgue*-integrable function. In particular, for $n \geq 2$, every orbit $O := \Phi(\mathbb{R}, \varphi) \subset \mathbb{T}^n$ is a Lebesgue-measurable subset of measure 0, but $\mathbb{1}_O^*(\varphi) = 1$.

[3] This means we assume that the Lebesgue measures of the interior of U and of the closure of U are equal.

- Otherwise there would exist an ε-neighborhood $U \subset \mathbb{T}^n$ of a point for which $\varphi(\mathbb{R}) \cap U = \emptyset$, hence $\langle \mathbb{1}_U \rangle > \mathbb{1}_U^*\big(\varphi(0)\big) = 0$. □

15.9 Remark (Unique Ergodicity)
From Theorem 9.14 by Koopman, it follows by an approximation argument that the conditionally periodic motion for independent frequencies is ergodic with respect to the Haar measure. From ergodicity alone, one could conclude with Birkhoff's ergodic theorem 9.32 that the time average equals the space average for λ^n-almost all initial values $\varphi_0 \in \mathbb{T}^n$.

Weyl's theorem is however stronger since it is valid for *all* initial conditions. This distinction is important, because exceptional sets of initial conditions would prevent the Cesàro-means in time, namely $T \mapsto T^{-1}\int_0^T \mathbb{1}_U\big(\varphi(t,\varphi_0)\big)\,\mathrm{d}t$, from converging to the space average uniformly in φ_0.

The underlying strong property of the dynamics to have only *one* invariant Borel probability measure is called *unique ergodicity*. In particular, such measures are then ergodic (why?). ♢

As a converse to Theorem 15.6, rational independence is necessary for the ergodicity of the conditionally periodic flow:

15.10 Lemma *If there exists a lattice vector $k \in \mathbb{Z}^n \backslash \{0\}$ with $\langle k, \omega \rangle = 0$, then there exists a continuous $f : \mathbb{T}^n \to \mathbb{C}$ whose time average f^* is not constant.*

Proof: Let $f(\varphi) := \exp(i\langle k, \varphi \rangle)$. Then the time average $f^*(\varphi)$ exists, because

$$\lim_{T\to\infty} \tfrac{1}{T}\int_0^T f(\varphi + \omega t)\,\mathrm{d}t = \lim_{T\to\infty} \tfrac{1}{T}\int_0^T \exp(i\langle k, \varphi + \omega t\rangle)\,\mathrm{d}t$$
$$= \tfrac{1}{T}\int_0^T \exp(i\langle k, \varphi \rangle)\,\mathrm{d}t = f(\varphi)\,.$$

So the non-constant function f is equal to its own time average f^*. □

Proof of Theorem 15.6 (Weyl's Theorem):

- The space average exists by the integrability of f.
- In order to prove the existence of the time average and to show, in the case of rational independence, that it equals the space average, we begin with simple functions f:

1. If f is constant, then time and space average are equal to this constant.
   If $f(\varphi) := \exp(i\langle k, \varphi\rangle)$ for some $k \in \mathbb{Z}^n \backslash \{0\}$, and $\langle k, \omega \rangle = 0$, then the time average exists according to the proof of Lemma 15.10.
   If $\langle k, \omega \rangle \neq 0$, then the time average exists as well, and in this case, it is equal to the space average:

$$f^*(\varphi) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \exp(i\langle k, \varphi + \omega t\rangle)\,\mathrm{d}t$$
$$= \exp(i\langle k, \varphi\rangle) \lim_{T\to\infty} \frac{1}{T}\,\frac{e^{i\langle k,\varphi\rangle T} - 1}{i\langle k, \varphi\rangle} = 0\,.$$

By linearity of the integral, the claim of the theorem follows for the finite linear combinations, namely the trigonometric polynomials.

2. Now we consider continuous functions $f : \mathbb{T}^n \to \mathbb{C}$. By linearity of the integral, we may assume f to be real valued with no loss of generality. By the Weierstrass approximation theorem, there exists, for every $\varepsilon > 0$, a trigonometric polynomial $p : \mathbb{T}^n \to \mathbb{C}$ with $|f - p| < \frac{1}{2}\varepsilon$. By replacing p with $\frac{1}{2}(p + \bar{p})$, we can ascertain that p is real valued, too. Then we conclude

$$\begin{aligned}&\limsup_{T\to\infty} \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t - \liminf_{T\to\infty} \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t \\ &\quad = \left(\limsup_{T\to\infty} \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t - p^*(\varphi)\right) - \left(\liminf_{T\to\infty} \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t - p^*(\varphi)\right) \\ &\quad \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.\end{aligned}$$

Hence the time limit exists. The equality of both averages for independent $\omega \in \mathbb{R}^n$ also carries over from trigonometric polynomials to continuous functions.

3. We finally arrive at the bounded Riemann integrable functions $f : \mathbb{T}^n \to \mathbb{R}$. While we cannot approximate them by continuous functions pointwise, we can do it in the integral norm.
By definition of the Riemann integral of f, for every $\varepsilon > 0$, there are step functions $g_1 \le f \le g_2$ with $(2\pi)^{-n} \int_{\mathbb{T}^n} (g_2 - g_1)\,\mathrm{d}\varphi < \frac{1}{4}\varepsilon$.
Moreover, there are continuous functions $f_1 \le g_1$ and $f_2 \ge g_2$ with

$$(2\pi)^{-n} \int_{\mathbb{T}^n} |f_k - g_k|\,\mathrm{d}\varphi < \tfrac{1}{4}\varepsilon\,.$$

This follows from the corresponding statement for the characteristic functions $g = 1\!\!1_Q$ of a rectangular box Q: Just set, for $\varphi \in \mathbb{T}^n$,

$$f_1(\varphi) := \min\bigl(c\ \mathrm{dist}(\varphi, \mathbb{T}^n - Q), 1\bigr) \quad , \quad f_2(\varphi) := \max\bigl(1 - c\ \mathrm{dist}(\varphi, Q), 0\bigr)$$

with $c > 0$. Then $f_1 \le g \le f_2$. Now choose c sufficiently large.
For continuous f_k, we have shown the existence of the time average f_k^*. We assume that the frequency vector $\omega \in \mathbb{R}^n$ is rationally independent. Thus $f_k^* = \langle f_k \rangle$. But as $f_1 < f < f_2$, we have $|\langle f \rangle - \langle f_k \rangle| < \varepsilon$. Therefore, the time average of f exists as well, and it equals $\langle f \rangle$. □

Excursion: The Virial Theorem

Space and time averages of particular phase space functions can be calculated in much more general situations. To this end, we consider first a completely general smooth Hamilton function $H : \mathbb{R}^{2n} \to \mathbb{R}$, for which the energy shell

$$\Sigma_E := \{x \in \mathbb{R}^{2n} \mid H(x) = E\}$$

for some value E of the energy is compact.
For an arbitrary function $f \in C^\infty(\mathbb{R}^{2n}, \mathbb{R})$, the time average of

$$\{f, H\} \in C^\infty\big(\mathbb{R}^{2n}, \mathbb{R}\big)$$

on Σ_E is then zero, because the flow Φ generated by H satisfies $\{f, H\} = \frac{\mathrm{d}}{\mathrm{d}t} f \circ \Phi_t|_{t=0}$, and therefore the time average for $x \in \Sigma_E$ can be calculated as

$$\{f, H\}^*(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \{f, H\} \circ \Phi_t(x)\, \mathrm{d}t = \lim_{T\to\infty} \frac{f \circ \Phi_T(x) - f(x)}{T} = 0\,,$$

using the boundedness of $f|_{\Sigma_E}$.

15.11 Example (Virial Theorem)
We consider the case of a Hamilton function

$$H(p, q) := T(p) + V(q) \quad \text{with kinetic energy} \quad T(p) := \tfrac{1}{2}\|p\|^2$$

(and always assume the energy shell Σ_E to be compact).

1. For the components of the momentum, $f(p, q) := p_i,\ i = 1, \ldots, n$, one obtains: The time average of the force $-\nabla V(q)$ acting on the mass point vanishes.
2. The analogous statement for the position components $q_i,\ i = 1, \ldots, n$ is: The time average of the momentum (which in the present case is just the velocity) is zero.
In contrast, for relativistic motion with its kinetic energy $T(p) := \sqrt{1 + \|p\|^2}$, only the time average of velocity vanishes, but not the time average of the momentum.
3. Returning to the nonrelativistic case ($T(p) = \frac{1}{2}\|p\|^2$), the Hamilton function for dilation $f(p, q) := \langle p, q\rangle$ satisfies

$$\{f, H\}(p, q) = 2\,T(p) - \langle q, \nabla V(q)\rangle\,.$$

At first sight, the second term lacks an intuitive interpretation. But if we choose the potential to be a homogenous polynomial of degree k in the q_i (with even k), we obtain (using multi-index notation)

$$V(q) = \sum_{\alpha\in\mathbb{N}_0^n, |\alpha|=k} c_\alpha q^\alpha \quad , \quad q_l \frac{\partial}{\partial q_l} V(q) = \sum_{\alpha\in\mathbb{N}_0^n, |\alpha|=k} \alpha_l c_\alpha q^\alpha\,,$$

and thus $\langle q, \nabla V(q)\rangle = k\,V(q)$. So the time averages of potential and kinetic energy (which exist almost everywhere by Birkhoff's ergodic theorem 9.32) satisfy

$$k\,V^*(x) = 2\,T^*(x)\,,$$

or

$$T^*(x) = \frac{k}{k+2} H(x) \quad , \quad V^*(x) = \frac{2}{k+2} H(x) \,.$$

One can argue analogously in the case of a homogenous central potential $V(q) = \|q\|^C$, in particular for the Kepler motion $C = -1$. In this case one has $T^*(x) = -H(x)$ and $V^*(x) = 2H(x) < 0$. (Note that T^* and V^* depend on all phase space variables!) ♢

The virial theorem is of significance in statistical mechanics. Namely, there one asks the question how the total energy is distributed across the different degrees of freedom.

15.12 Exercise (Virial Theorem for Space Averages) Let E be a regular value of the Hamilton function $H \in C^2(\mathbb{R}^n_p \times \mathbb{R}^n_q, \mathbb{R})$. Let λ_E denote the Liouville measure on the energy shell Σ_E, which is assumed to be compact, normalized as a probability measure: $\lambda_E(\Sigma_E) = 1$. We set $\langle g \rangle_E := \int_{\Sigma_E} g \, \mathrm{d}\lambda_E$.

(a) Prove the virial theorem for space averages: For arbitrary phase space functions $f \in C^2\big(\mathbb{R}^n_p \times \mathbb{R}^n_q, \mathbb{R}\big)$, one has $\langle \{f, H\} \rangle_E = 0$.
(b) Carry over the Examples 15.11 to their variants for space averages.
(c) In relation to the equation of state for a real gas with $N \in \mathbb{N}$ particles in the configuration space $\mathbb{R}^n_q = \mathbb{R}^{3N}$, we denote their positions as $q = (q_1, \ldots, q_N)$. Let the Hamilton function be

$$H(p,q) := \sum_{j=1}^{N} \Big(\tfrac{1}{2} \|p_j\|^2 + U_j(q_j) + \sum_{k=j+1}^{N} W_{j,k}(q_j - q_k) \Big)$$

$\big((p,q) \in \mathbb{R}^n_p \times \mathbb{R}^n_q\big)$, with the potentials $U_j, W_{j,k} : \mathbb{R}^n \to \mathbb{R}$.
What does the virial theorem say for $f(p,q) := \langle p, q \rangle$?
(d) Let the container in which the particles are confined be the compact closure $G \subset \mathbb{R}^3$ of a domain with smooth boundary. Let the Hamilton function be more simply

$$H_\lambda(p,q) := \textstyle\sum_{j=1}^{N} \big(\tfrac{1}{2} \|p_j\|^2 + \lambda U(q_j) \big) \qquad \big((p,q) \in \mathbb{R}^n_p \times \mathbb{R}^n_q,\ \lambda > 0\big),$$

with the container potential

$$U(q) := \mathrm{dist}(q, G)^2 \quad \text{, for} \quad \mathrm{dist}(q,G) = \min\{\|q - Q\| \mid Q \in G\} \,.$$

Show how we obtain, in the limit $\lambda \to \infty$, the equation of state of an ideal gas $PV = \frac{2}{3} \langle T \rangle_E$ with the pressure P, the volume V of G, and kinetic energy T.
Hints: Investigate first the differentiability properties of U, in particular near the boundary ∂G. Recall the definition of pressure in physics and use the Stokes theorem B.39. ♢

15.2 Perturbation Theory for One Angle Variable

Before considering Hamiltonian systems, we study the perturbation theory in a case in which the averaging principle can be applied particularly well. The perturbed differential equation on the phase space $G \times S^1$, with $G \subseteq \mathbb{R}^m$ open, is assumed to be

$$\dot{I} = \varepsilon g(I, \varphi) \quad , \quad \dot{\varphi} = \omega(I) + \varepsilon f(I, \varphi)\,, \tag{15.2.1}$$

with $\omega, f \in C^1(G \times S^1, \mathbb{R})$ and $g \in C^1(G \times S^1, \mathbb{R}^m)$. So the averaged system is

$$\dot{J} = \varepsilon \langle g\rangle(J)\,. \tag{15.2.2}$$

We denote the unique solution of the perturbed system as $t \mapsto \big(I_\varepsilon(t), \varphi_\varepsilon(t)\big)$, and the solution of the averaged system as $t \mapsto J_\varepsilon(t)$, both for initial condition $I(0) = J(0) = I_0$, $\varphi(0) = \varphi_0$. As the variation

$$\tilde{g}(I, \varphi) := g(I, \varphi) - \langle g\rangle\,(I) \tag{15.2.3}$$

of g does not vanish in general, one might expect that it is only for times t of order $\mathcal{O}(\varepsilon^0) = \mathcal{O}(1)$ that the averaged and the original system differ only by order $\mathcal{O}(\varepsilon)$. This is however not the case.

15.13 Theorem (Perturbation Theory for One Angle Variable, 1st Order)

1. *Assume the frequency function ω in (15.2.1) does not take the value zero.*
2. *Assume that for initial value $(I_0, \varphi_0) \in G \times S^1$ and a compact subset $G_k \subset G$, the initial value problem of the averaged system (15.2.2) has solutions*

$$J_\varepsilon : [0, 1/\varepsilon] \to G_k \qquad \big(0 < \varepsilon \le \varepsilon_0\big).$$

Then for small ε, the true and the averaged solution stay closely together for a long time:

$$\sup_{0\le t\le 1/\varepsilon} \|I_\varepsilon(t) - J_\varepsilon(t)\| = \mathcal{O}(\varepsilon) \qquad \big(0 < \varepsilon \le \varepsilon_0\big).$$

Proof: We suppress the indices ε in the proof.

- The idea of the proof is to introduce new, slowly varying coordinates

$$\tilde{I}(I, \varphi) := I + \varepsilon K(I, \varphi) \tag{15.2.4}$$

in which the angle dependent perturbation is only of order ε^2. A perturbation of this size can then be integrated over a time interval of length $1/\varepsilon$, and we obtain a maximal deviation from the unperturbed system in the order ε. To determine the change of coordinates, we plug the ansatz (15.2.4) into the differential equation (15.2.1) and obtain

$$\dot{\tilde{I}} = \dot{I} + \varepsilon\big[\mathrm{D}_1 K(I,\varphi)\dot{I} + \mathrm{D}_2 K(I,\varphi)\dot{\varphi}\big] = \varepsilon\,[g(I,\varphi) + \mathrm{D}_2 K(I,\varphi)\omega(I)] + \varepsilon^2 R(I,\varphi) \tag{15.2.5}$$

with the remainder term

$$R(I,\varphi) \;:=\; \mathrm{D}_1 K(I,\varphi) g(I,\varphi) \;+\; \mathrm{D}_2 K(I,\varphi) f(I,\varphi)\,.$$

- We now choose

$$K(I,\varphi) := -\frac{1}{\omega(I)}\int_0^{\varphi} \tilde{g}(I,\psi)\,\mathrm{d}\psi \qquad \big((I,\varphi)\in G_k\times S^1\big), \tag{15.2.6}$$

  with $\tilde{g}$ from (15.2.3) in order to make the parenthesis in (15.2.5) independent of φ. This is possible because of the hypothesis $\omega(I)\neq 0$. Since $\langle\tilde{g}\rangle(I)=0$, it is indeed true that K is 2π-periodic in φ.
- As the compact set G_k has a positive distance to the closed set $\mathbb{R}^m\backslash G$, there is even a closed δ-neighborhood of G_k in G. We will again denote this likewise compact neighborhood by G_k. Since it is compact, both $K\in C^1(G_k\times S^1,\mathbb{R}^m)$ and its derivative are bounded. Likewise, $R\in C^0(G_k\times S^1,\mathbb{R}^m)$ is bounded. For (15.2.6), the parenthesis in (15.2.5) equals $\langle g\rangle(I)$.
- We further conclude that for small ε, we can solve (15.2.4) for I, and that the original action $I(\tilde{I},\varphi)$ and its derivative are again bounded. Its domain $\big\{\big(\tilde{I}(I,\varphi),\varphi\big)\mid (I,\varphi)\in G_k\times S^1\big\}$ is contained in $G\times S^1$ for small $\varepsilon>0$. We compare the differential equation

$$\dot{\tilde{I}} = \varepsilon\langle g\rangle(I) + \varepsilon^2 R(I,\varphi) = \varepsilon\langle g\rangle(\tilde{I}) + \varepsilon^2 R(I,\varphi) + \varepsilon\Big(\langle g\rangle(I) - \langle g\rangle(\tilde{I})\Big)$$

of $\tilde{I}$ with the one for the averaged variable. Letting $\Delta I := \tilde{I} - J$ and using (15.2.4), the initial value is $\Delta I(0) = \tilde{I}(0) - I(0) = \mathcal{O}(\varepsilon)$, and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Delta I &= \varepsilon\Big(\langle g\rangle(\tilde{I}) - \langle g\rangle(J)\Big) + \varepsilon^2 R(I,\varphi) + \varepsilon\Big(\langle g\rangle(I) - \langle g\rangle(\tilde{I})\Big)\\ &= \varepsilon\,\mathrm{D}\langle g\rangle(J)\cdot\Delta I + \varepsilon^2 R(I,\varphi) + \varepsilon\Big(\langle g\rangle(I) - \langle g\rangle(\tilde{I})\Big) + \\ &\quad\ \varepsilon\Big(\langle g\rangle(\tilde{I}) - \langle g\rangle(J) - \mathrm{D}\langle g\rangle(J)\cdot\Delta I\Big)\,.\end{aligned}$$

As $g\in C^1(G\times S^1,\mathbb{R}^m)$, the last term is of order $o(\varepsilon^2)$ by Taylor's formula, if $\Delta I = \mathcal{O}(\varepsilon)$. We make this assumption for the time interval $[0,1/\varepsilon]$ and check the consistency of this assumption.

- ΔI satisfies the integral equation

$$\Delta I(t) = \Delta I(0) + \int_0^t \Big[\varepsilon\mathrm{D}\langle\tilde{g}\rangle\big(J(s)\big)\cdot\Delta I(s) + \tilde{R}(s,\varepsilon)\Big]\,\mathrm{d}s\,, \tag{15.2.7}$$

where by hypothesis,

$$\tilde{R}(s,\varepsilon) := \varepsilon^2 R\Big(I\big(\tilde{I}(s),\varphi(s)\big),\varphi(s)\Big) + \varepsilon\Big(\langle g\rangle\,\big(I\big(\tilde{I}(s),\varphi(s)\big),\varphi(s)\big) - \langle g\rangle\,\big(\tilde{I}(s)\big)\Big)$$
$$+\varepsilon\,\Big(\langle g\rangle\,\big(\tilde{I}(s)\big) - \langle g\rangle\,\big(J(s)\big) - \mathrm{D}\,\langle g\rangle\,\big(J(s)\big)\cdot\Delta I(s)\Big)$$

is of the order $\mathcal{O}(\varepsilon^2)$. We estimate ΔI by means of (15.2.7), and to this end we use the Gronwall lemma (Theorem 3.42) with

$$F := \|\Delta I\| \quad , \quad a := \|\Delta I(0)\| + \int_0^{1/\varepsilon} \Big\|\tilde{R}(s,\varepsilon)\Big\|\,\mathrm{d}s \quad \text{and} \quad G := \varepsilon\,\|\mathrm{D}\,\langle g\rangle\|\,.$$

If the solution $t \mapsto I(t)$ remains in G_k (which is guaranteed for $F(t) \le \delta$), the Gronwall inequality (3.6.3) takes the form

$$F(t) \;\le\; c_1\varepsilon\; \exp(c_2\varepsilon\, t) \qquad (0 \le t \le 1/\varepsilon).$$

This would be consistent with our assumption $\Delta I = \mathcal{O}(\varepsilon)$.

We can choose the maximum size ε_0 of the perturbation so small that

$$c_1\varepsilon_0\; \exp(c_2) < \tfrac{1}{2}\delta \quad \text{and} \quad \varepsilon_0 \sup_{(I,\varphi)\in G_k\times S^1} \|K(I,\varphi)\| < \tfrac{1}{2}\delta\,,$$

hence $\|I(t) - J(t)\| < \delta$, and thus for $0 \le t \le 1/\varepsilon$, indeed $I(t) \in G_k$. □

The procedure can be iterated if ω, f and g have higher differentiability, and under favorable hypotheses yields in nth order a control of the solution in the time interval $[0, 1/\varepsilon^n]$.

15.3 Hamiltonian Perturbation Theory of First Order

"Whether I apply mathematics to a couple of lumps of dirt, which we call planets, or to purely arithmetical problems, it's the same, only the latter have a yet higher attraction to me." (Carl Friedrich Gauss)[4]

We consider a Hamiltonian system with Hamilton function

$$H_\varepsilon(I,\varphi) := H_0(I) + \varepsilon H_1(I,\varphi) \tag{15.3.1}$$

on the phase space $G\times\mathbb{T}^n$, where $G \subset \mathbb{R}^n$ is open and bounded, with the perturbation parameter $|\varepsilon| < \varepsilon_0$. So the differential equations are

$$\dot{I} = -\varepsilon\,\mathrm{D}_2 H_1(I,\varphi) \quad , \quad \dot{\varphi} = \omega(I) + \varepsilon\,\mathrm{D}_1 H_1(I,\varphi)$$

[4]Quoted after W. Sartorius v. Waltershausen: Gauss zum Gedächtniss. 1856, Reprint 1965, page 101/102, (and translated from this source).

with the frequency vector

$$\omega := \mathrm{D}H_0 : G \to \mathbb{R}^n\,. \tag{15.3.2}$$

The averaged system is trivial: $\dot{J} = 0$. For $n = 1$ degree of freedom, we can apply Theorem 15.13 from the last section, provided the rotation frequency does not vanish. We then obtain the conclusion

$$I(t) = I(0) + \mathcal{O}(\varepsilon) \qquad (0 \le t \le 1/\varepsilon).$$

For $n > 1$ degrees of freedom, we are looking for a canonical transformation

$$T_\varepsilon : (I, \varphi) \longmapsto (\tilde{I}, \tilde{\varphi})$$

that eliminates the angle dependence of the differential equation up to a term of order ε^2. To this end, we use the method of a generating function and make the ansatz

$$I = \tilde{I} + \varepsilon\, \mathrm{D}_2 S(\tilde{I}, \varphi) \quad , \quad \tilde{\varphi} = \varphi + \varepsilon\, \mathrm{D}_1 S(\tilde{I}, \varphi)\,. \tag{15.3.3}$$

Plugging this into the Hamiltonian and using $K_\varepsilon \circ T_\varepsilon = H_\varepsilon$ yields formally

$$\begin{aligned} K_\varepsilon(\tilde{I}, \tilde{\varphi}) &= H_\varepsilon\Big(\tilde{I} + \varepsilon\, \mathrm{D}_2 S\big(\tilde{I}, \varphi(\tilde{I}, \tilde{\varphi})\big)\,,\ \tilde{\varphi} - \varepsilon\, \mathrm{D}_1 S\big(\tilde{I}, \varphi(\tilde{I}, \tilde{\varphi})\big)\Big) \\ &= H_0(\tilde{I}) + \varepsilon\Big[\big\langle \mathrm{D}H_0(\tilde{I}), \mathrm{D}_2 S(\tilde{I}, \tilde{\varphi})\big\rangle + H_1(\tilde{I}, \tilde{\varphi})\Big] + \mathcal{O}(\varepsilon^2)\,. \end{aligned} \tag{15.3.4}$$

Thus S would have to be chosen in such a way that

$$\big\langle \omega(\tilde{I}), \mathrm{D}_2 S(\tilde{I}, \tilde{\varphi})\big\rangle + H_1(\tilde{I}, \tilde{\varphi})$$

becomes a function of $\tilde{I}$ only, not of the angles. To this end, one uses the Fourier transform and thus writes (initially only in the sense of formal series)

$$H_1(\tilde{I}, \tilde{\varphi}) = \sum_{\ell \in \mathbb{Z}^n} h_\ell(\tilde{I}) \exp(i\,\langle \ell, \tilde{\varphi}\rangle) \quad , \quad S(\tilde{I}, \tilde{\varphi}) = \sum_{\ell \in \mathbb{Z}^n} S_\ell(\tilde{I}) \exp(i\,\langle \ell, \tilde{\varphi}\rangle)\,. \tag{15.3.5}$$

Since $\big\langle \mathrm{D}_2 S(\tilde{I}, \tilde{\varphi}), \omega(\tilde{I})\big\rangle = i \sum_{\ell \in \mathbb{Z}^n} S_\ell(\tilde{I})\, \langle \ell, \omega(\tilde{I})\rangle\ \exp(i\,\langle \ell, \tilde{\varphi}\rangle)$, one obtains as conditions the equations

$$i S_\ell(\tilde{I})\ \langle \ell, \omega(\tilde{I})\rangle + h_\ell(\tilde{I}) = 0 \qquad \big(\ell \in \mathbb{Z}^n \backslash \{0\}\big). \tag{15.3.6}$$

In general, these equations cannot be solved, because

- for $n > 1$ degrees of freedom, we can always, by an arbitrarily small change of the frequency vector $\omega \in \mathbb{R}^n$, make the scalar product $\langle \ell, \omega\rangle$ vanish for some appropriate $\ell \in \mathbb{Z}^n \backslash \{0\}$; and

- if the mapping $\mathrm{D}\omega : G \to \mathrm{Mat}(n, \mathbb{R})$ has maximum rank n, we can achieve such a change of ω by a variation of $\tilde{I}$.

However, if we assume that for a *fixed* $\hat{I} \in G$, the components of the frequency vector $\omega(\hat{I}) \in \mathbb{R}^n$ are *rationally independent*, i.e.,

$$\langle \ell, \omega(\hat{I}) \rangle \neq 0 \qquad \big(\ell \in \mathbb{Z}^n \backslash \{0\}\big),$$

then we can at least solve the equations (15.3.6) at the point $\tilde{I} = \hat{I}$ by choosing the Fourier coefficients (independent of $\tilde{I}$) as

$$S_\ell(\tilde{I}) := -i \frac{h_\ell(\hat{I})}{\langle \ell, \omega(\hat{I}) \rangle} \quad \big(\ell \in \mathbb{Z}^n \backslash \{0\}\big) \quad \text{and} \quad S_0(\tilde{I}) := 0 \,. \tag{15.3.7}$$

However, there appears *the problem of small denominators* in (15.3.7). While the Fourier coefficients S_ℓ are defined, it is not clear whether the Fourier series (15.3.5) of S converges.

In view of this problem, we will strengthen the condition of independence and require that $\omega(\hat{I}) \in \mathbb{R}^n$ is *diophantine*, i.e., for appropriate $\gamma > 0$ and $\tau > 0$, we require ω to lie in the set

$$\boxed{\Omega_{\gamma,\tau} := \left\{ \hat{\omega} \in \mathbb{R}^n \mid \forall \ell \in \mathbb{Z}^n \backslash \{0\} : \left| \langle \ell, \hat{\omega} \rangle \right| \geq \gamma |\ell|^{-\tau} \right\}.} \tag{15.3.8}$$

We postpone the discussion whether any vector $\hat{\omega} \in \mathbb{R}^n$ at all satisfies such a condition (see Lemma 15.18).

Moreover we require, in order to improve the odds of convergence for (15.3.5), that the Fourier coefficients h_ℓ decay rapidly. This is the case if H_1 has sufficient differentiability:

15.14 Lemma (Differentiability and Fourier Coefficients)
For a function $g \in C^k(\mathbb{T}^n, \mathbb{R})$ on the torus $\mathbb{T}^n$, with the Fourier representation

$$g(\varphi) = \textstyle\sum_{\ell \in \mathbb{Z}^n} g_\ell \, \exp(i \, \langle \ell, \varphi \rangle) \,,$$

the Fourier coefficients $g_\ell \in \mathbb{C}$ are of the order

$$|g_\ell| = \mathcal{O}\big(|\ell|^{-k}\big) \,. \tag{15.3.9}$$

Proof: For a multi-index $\alpha \in \mathbb{N}_0^n$ of length $|\alpha| = \sum_{j=1}^n \alpha_j \leq k$, we have

$$\mathrm{D}^\alpha g(\varphi) = i^{|\alpha|} \textstyle\sum_{\ell \in \mathbb{Z}^n} g_\ell \, \ell^\alpha \, \exp(i \, \langle \ell, \varphi \rangle) \,. \tag{15.3.10}$$

This is proved by integration by parts. Now let $\ell \in \mathbb{Z}^n \backslash \{0\}$ and ℓ_j a component of ℓ with maximum absolute value, so in particular $|\ell_j| \geq |\ell|/n$. By an inverse Fourier transform of the kth derivative with respect to φ_j, namely

$$g_\ell = \frac{1}{(i\ell_j)^k} \int_{\mathbb{T}^n} \exp(-i\,\langle \ell, \varphi\rangle)\,\partial^k_{\varphi_j} g(\varphi)\,\frac{d\varphi}{(2\pi)^n}\,,$$

we prove (15.3.9), because $\max_j \sup_\varphi |\partial^k_{\varphi_j} g(\varphi)| < \infty$. □

These types of statements are called *Paley-Wiener estimates*. Conversely, how does the order of differentiability of a function depend on decay properties of its Fourier coefficients?

15.15 Lemma (Differentiability of Fourier Series) *If $c : \mathbb{Z}^n \to \mathbb{C}$ is of the order $c(\ell) = \mathcal{O}(|\ell|^{-k})$ for some real number $k > n + r$, $r \in \mathbb{N}_0$, then the function f defined by*

$$f(\varphi) := \sum_{\ell\in\mathbb{Z}^n} c(\ell)\,\exp(i\,\langle \ell, \varphi\rangle) \qquad (\varphi \in \mathbb{T}^n) \tag{15.3.11}$$

is in $C^r(\mathbb{T}^n, \mathbb{C})$.

Proof: By (15.3.10), it suffices to prove the claim for $r = 0$, i.e., to prove the continuity of f. Thus there exists $C > 0$ with $|c(\ell)| \le C|\ell|^{-k}$. The series (15.3.11) converges uniformly because

$$\sum_{\ell\in\mathbb{Z}^n} |c(\ell)| \le |c(0)| + C \sum_{\ell\in\mathbb{Z}^n\setminus\{0\}} |\ell|^{-k} < \infty\,. \tag{15.3.12}$$

The sum can be controlled by comparison with the integral $\int_R \|x\|^{-k}\,dx = c_n \int_1^\infty r^{n-1-k}\,dr$ for the domain of integration $R := \{x \in \mathbb{R}^n \mid \|x\| \ge 1\}$. For $\varepsilon > 0$, let N be chosen in such a way that

$$\sum_{\ell\in\mathbb{Z}^n,\, |\ell|>N} |c(\ell)| \le \varepsilon/3\,, \tag{15.3.13}$$

and let f_N be the partial sum

$$f_N(\varphi) := \sum_{\ell\in\mathbb{Z}^n,\, |\ell|\le N} c(\ell)\exp\bigl(i\,\langle \ell, \varphi\rangle\bigr)\,.$$

Then its continuity carries over to f by means of an $\varepsilon/3$-argument: For all $\varphi, \psi \in \mathbb{T}^n$,

$$\begin{aligned} |f(\varphi) - f(\psi)| &\le |f(\varphi) - f_N(\varphi)| + |f_N(\varphi)| - f_N(\psi)| + |f_N(\psi) - f(\psi)| \\ &\le \frac{\varepsilon}{3} + |f_N(\varphi)| - f_N(\psi)| + \frac{\varepsilon}{3}\,. \end{aligned}$$

On the other hand, if $|\varphi - \psi| < \delta := \frac{\varepsilon}{3CM}$ with $M := \sum_{\ell\in\mathbb{Z}^n\setminus\{0\},\, |\ell|\le N} |\ell|^{1-k}$, then

$$
\begin{aligned}
|f_N(\varphi) - f_N(\psi)| &= \left| \sum_{\ell \in \mathbb{Z}^n,\, |\ell| \le N} c(\ell) \big(\exp(i\,\langle \ell, \varphi\rangle) - \exp(i\,\langle \ell, \psi\rangle)\big) \right| \\
&\le C \sum_{\ell \in \mathbb{Z}^n \setminus \{0\},\, |\ell| \le N} |\ell|^{-k} \big| 1 - \exp(i\,\langle \ell, \psi - \varphi\rangle) \big| \\
&\le C \sum_{\ell \in \mathbb{Z}^n \setminus \{0\},\, |\ell| \le N} |\ell|^{-k}\, |\langle \ell, \psi - \varphi\rangle| < C \sum_{\ell \in \mathbb{Z}^n \setminus \{0\},\, |\ell| \le N} |\ell|^{-k} |\ell| \delta \le \frac{\varepsilon}{3}\,,
\end{aligned}
$$

hence[5] $|f(\varphi) - f(\psi)| < \varepsilon$. □

15.16 Corollary *For*[6] $H_1 \in C^k_{\mathrm{b}}(G \times \mathbb{T}^n, \mathbb{R})$ *with* $k > n + \tau$, $\tau \notin \mathbb{N}$ *and an action* $\hat{I} \in G$ *with* $\omega(\hat{I})$ *satisfying the diophantine condition (15.3.8), the generating function* S *with Fourier coefficients (15.3.7) satisfies*[7]

$$S \in C^{k-\lceil \tau \rceil - n}_{\mathrm{b}}\big(G \times \mathbb{T}^n, \mathbb{R}\big)\,.$$

Moreover, there exists a $C > 0$ *such that*

$$\sup\left\{ |\partial^\alpha S| \mid \hat{I} \in G,\ \omega(\hat{I}) \in \Omega_{\gamma,\tau} \right\} \le \tfrac{C}{\gamma} \qquad \big(|\alpha| \le k - n - \tau\big). \tag{15.3.14}$$

Proof: Applying Lemma 15.14 to H_1 leads to the Fourier coefficients (15.3.7) satisfying

$$|S_\ell| = \left| \tfrac{h_\ell(\hat{I})}{\langle \ell, \omega(\hat{I})\rangle} \right| \le \tfrac{C|\ell|^{\tau-k}}{\gamma} \qquad \big(\ell \in \mathbb{Z}^n \setminus \{0\}\big). \tag{15.3.15}$$

Thus by Lemma 15.15, it follows that $S \in C^{k-\lceil\tau\rceil-n}_{\mathrm{b}}\big(G \times \mathbb{T}^n, \mathbb{C}\big)$. S is real valued because the Fourier coefficients in (15.3.7) have the symmetry $\overline{S_\ell} = S_{-\ell}$.

The constants in (15.3.15) carry over to (15.3.14), due to the linearity of the Fourier transform and the derivative. □

15.17 Remark (**Cohomological Equation**)
After having done the analytical work, we should reassess it from a conceptual point of view. We considered just one action, $\hat{I} \in G$, for which $\hat{\omega} := \omega(\hat{I}) \in \mathbb{R}^n$ (with ω defined in (15.3.2)) is diophantine ($\hat{\omega} \in \Omega_{\gamma,\tau}$). Accordingly, the generating function $S : G \times \mathbb{T}^n \to \mathbb{C}$ did not depend on $\tilde{I} \in G$. We thus interpret it just as a function $S : \mathbb{T}^n \to \mathbb{C}$. This function is given in terms of its Fourier coefficients (15.3.7), which in turn depend on the perturbation $H_1(\hat{I}, \cdot) : \mathbb{T}^n \to \mathbb{C}$. Again, we omit the dependence on the action and write $H_1 : \mathbb{T}^n \to \mathbb{C}$. Then S satisfies the *cohomological equation*

[5] But one also notes that differentiability for $k \le n + 1$ does not carry over from f_N to f, because then M diverges as $N \to \infty$.

[6] $C^k_{\mathrm{b}}(U, \mathbb{R})$ denotes the vector space of those k times continuously differentiable functions on $U \subseteq \mathbb{R}^m$ whose partial derivatives up to this order are bounded. Here we are using angle coordinates on the torus $\mathbb{T}^n$.

[7] With the *ceiling* function $\lceil\cdot\rceil : \mathbb{R} \to \mathbb{Z}\,,\ x \mapsto \min\{z \in \mathbb{Z} \mid z \ge x\}$, see page 52.

$$\boxed{L_{\hat\omega} S = H_1 - \langle H_1 \rangle\,.}$$

Here $\hat\omega$ is considered as a constant vector field on the torus $\mathbb{T}^n$, and $L_{\hat\omega}$ is the Lie derivative (or directional derivative). The validity of the cohomological equation can be checked as follows, using (15.3.7):

$$L_{\hat\omega} S(\varphi) = \sum_{\ell\in\mathcal{L}\backslash\{0\}} \frac{-ih_\ell}{\langle \ell, \hat\omega\rangle} L_{\hat\omega} \exp(i\,\langle \ell, \varphi\rangle) = \sum_{\ell\in\mathcal{L}\backslash\{0\}} h_\ell \exp(i\,\langle \ell, \varphi\rangle) = H_1(\varphi) - \langle H_1\rangle\,.$$

Solving the cohomological equation for S amounts to inverting the linear operator $L_{\hat\omega}$. As $L_{\hat\omega}$ is a derivation, its inverse should be a kind of integral operator.
Indeed, for $n = 1$ and $\hat\omega \in \mathbb{R}\backslash\{0\}$ (and assuming $S(0) := 0$, since $\ker(L_{\hat\omega})$ consists of the constant functions),

$$(L_{\hat\omega})^{-1}(H_1 - \langle H_1\rangle)(\varphi) = \hat\omega^{-1} \int_0^\varphi (H_1(\psi) - \langle H_1\rangle)\,\mathrm{d}\psi \qquad (\varphi \in \mathbb{R})$$

is a 2π-periodic function. However, for $n > 1$, we have seen that solvability crucially depends on the number-theoretical properties of $\hat\omega \in \mathbb{R}^n$.
We will meet the cohomological equation again in the context of KAM theory. $\Diamond$

The diophantine condition (15.3.8) is satisfied for Lebesgue-almost all $\omega \in \mathbb{R}^n$, provided $\tau > n - 1$ (but with a constant γ that depends on ω):

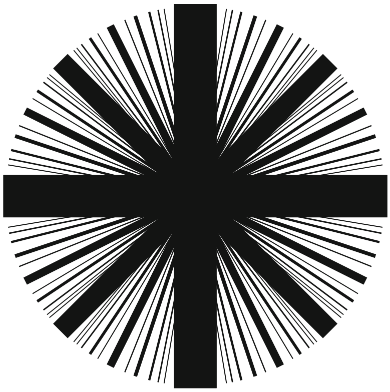


Figure 15.3.1 The diophantine set $B_{\gamma,\tau}$ defined in (15.3.16), in white

15.18 Lemma *For $\tau > n - 1$, the Lebesgue measure of the diophantine set*

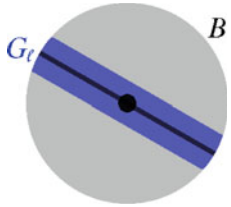
$$B_{\gamma,\tau} := \Omega_{\gamma,\tau} \cap B \tag{15.3.16}$$

in the ball $B := \{\omega \in \mathbb{R}^n \mid \|\omega\| \le 1\}$ is large for small values of γ: There exists $\alpha(\tau) < \infty$ such that

$$\lambda^n(B_{\gamma,\tau}) \ge \big(1 - \gamma\,\alpha(\tau)\big)\,\lambda^n(B)\,.$$

Proof: For the neighborhoods $G_\ell := \{\hat{\omega} \in B \mid |\langle \ell, \hat{\omega}\rangle| < \gamma|\ell|^{-\tau}\}$ of rational hypersurfaces, one has $B_{\gamma,\tau} = B \backslash \sum_{\ell\in\mathbb{Z}^n\backslash\{0\}} G_\ell$ (see figures. 15.3.1 and below), so that

$$\lambda^n(B_{\gamma,\tau}) \geq \lambda^n(B) - \sum_{\ell\in\mathbb{Z}^n\backslash\{0\}} \lambda^n(G_\ell)\,.$$



Now $\lambda^n(G_\ell) \leq 2v_{n-1}\gamma|\ell|^{-\tau-1}$, where v_k denotes the volume of the k-dimensional ball of radius 1 (see the figure). This implies (with an estimate for the sum as in the proof of (15.3.12)) that

$$\sum_{\ell\in\mathbb{Z}^n\backslash\{0\}} \lambda^n(G_\ell) \leq v_n\,\gamma\,\alpha(\tau) \quad \text{for} \quad \alpha(\tau) := 2\frac{v_{n-1}}{v_n} \sum_{\ell\in\mathbb{Z}^n\backslash\{0\}} |\ell|^{-\tau-1} < \infty\,. \qquad \square$$

15.19 Remark (Diophantine Set as a Cantor Set)
$\Omega_{\gamma,\tau}$ is the union of closed rays $\{s\,\hat{\omega} \mid \hat{\omega} \in \Omega_{\gamma,\tau},\ s \geq 1\}$.
This suggests to consider (and doing so is useful for the perturbation theory at constant energy, see Example 15.35) intersections of the form $\Omega_{\gamma,\tau} \cap S_R^{n-1}$ with spheres of radius $R > 0$.

These intersections are compact and totally disconnected. They are the union of a topological Cantor set (see page 489) and a countable set (where either constituent might be omitted). This follows from the theorem by Cantor and Bendixson, according to which a closed uncountable set can be decomposed into a countable set and a perfect set.[8] See BROER and SEVRYUK [BrSe], Chapter 4.1.2.

So we see here that the Lebesgue measure of a topological Cantor set in $\mathbb{R}^d$ does not have to be zero (which was however the case for the Cantor-$1/3$ set from Example 2.5). $\Diamond$

Let us gather our partial results so far:

15.20 Theorem (Hamiltonian Perturbation Theory of First Order)
If $H_0, H_1 \in C_b^{2n+3}(G\times\mathbb{T}^n, \mathbb{R})$ in (15.3.1), and if the frequencies $\omega = \mathrm{D}H_0 : G \to \mathbb{R}^n$ vary independently, i.e.,

$$\inf_{I\in G} |\det(\mathrm{D}\omega)(I)| > 0\,, \tag{15.3.17}$$

then there exist subsets $G_\gamma \subset G$ of asymptotically full measure, i.e., satisfying $\lim_{\gamma\searrow 0} \lambda^n(G_\gamma) = \lambda^n(G)$, for which the averaging principle applies in the following sense:

$$\sup_{x_0\in G_\gamma\times\mathbb{T}^n}\ \sup_{0\leq t\leq 1/\varepsilon} \|I_\varepsilon(t,x_0) - I_\varepsilon(0,x_0)\| = \mathcal{O}_\gamma(\varepsilon) \qquad \big(0 < \varepsilon \leq \varepsilon_0\big).$$

Proof: As a consequence of (15.3.17), $\omega : G \to \mathbb{R}^n$ is a local diffeomorphism.

[8] A perfect set is a set that is equal to the set of its cluster points.

- For $\tau := n - 1/2$, the set

$$G_\gamma := \left\{ I \in G \mid \mathrm{dist}(I, \mathbb{R}^n \backslash G) > \gamma\,,\ \omega(I) \in \Omega_{\gamma,\tau} \right\}$$

  of the nonresonant actions away from the boundary has full measure in the limit $\gamma \searrow 0$. This follows from Lemma 15.18 by means of (15.3.17).

- Now let $\hat{I} \in G_\gamma$. We consider the solutions for the initial conditions $x_0 \in \{\hat{I}\} \times \mathbb{T}^n$. Under the given hypotheses, the generating function S satisfies $S \in C^3(G \times \mathbb{T}^n, \mathbb{R})$ by Corollary 15.16, hence the estimate (15.3.4) holds with the constant term $\varepsilon h_1(\hat{I})$ (proportional to ε !) in an ε-neighborhood of $\hat{I}$. This means that by introducing the new coordinates $(\tilde{I}, \tilde{\varphi})$ (see (15.3.3)) the Hamilton function will be integrable up to error terms R_i of the order ε^2:

$$K_\varepsilon(\tilde{I}, \tilde{\varphi}) = H_0(\tilde{I}) + \varepsilon h_1(\hat{I}) + R_1(\tilde{I}, \tilde{\varphi}) + R_2(\tilde{I}, \tilde{\varphi})$$

  with

$$\begin{aligned} R_1(\tilde{I}, \tilde{\varphi}) := {} & \Big(H_0\big(\tilde{I} + \varepsilon \mathrm{D}_2 S(\tilde{I}, \tilde{\varphi})\big) - H_0(\tilde{I}) - \varepsilon \mathrm{D} H_0(\tilde{I}) \cdot \mathrm{D}_2 S\big(\tilde{I}, \tilde{\varphi}\big) \Big) \\ & + \varepsilon \Big(H_1\big(\tilde{I} + \varepsilon \mathrm{D}_2 S(\tilde{I}, \tilde{\varphi}), \tilde{\varphi}\big) - H_1\big(\tilde{I}, \tilde{\varphi}\big) \Big) = \mathcal{O}_\gamma\big(\varepsilon^2\big)\,, \end{aligned}$$

and the same estimate holds for the vector field X_{R_1}. The analogous statement applies to

$$R_2\big(\tilde{I}, \tilde{\varphi}\big) := \varepsilon \Big(\omega(\tilde{I}) \cdot \mathrm{D}_2 S(\tilde{I}, \tilde{\varphi}) - \omega(\hat{I}) \mathrm{D}_2 S(\hat{I}, \tilde{\varphi}) \Big) + \varepsilon \Big(H_1\big(\tilde{I}, \tilde{\varphi}\big) - H_1\big(\hat{I}, \tilde{\varphi}\big) \Big)\,.$$

- Integrating the Hamiltonian differential equations of K_ε by means of the Gronwall inequality (Theorem 3.42), as in the proof of Theorem 15.13, yields the result, provided the trajectories $t \mapsto I_\varepsilon(t, x_0)$ with initial values $x_0 \in G_\gamma \times \mathbb{T}^n$ remain in G during the time interval $t \in [0, 1/\varepsilon]$. This can be achieved by making ε smaller if necessary. □

15.21 Remark (Application to the Orbit of the Earth)
The masses of the giant planets Jupiter and Saturn are in the ballpark of $\varepsilon := 1/1000$ of the solar mass. So under their perturbations, the semiaxes of the orbit of the earth could change significantly after some $1/\varepsilon = 1000$ years.

If the hypotheses of Theorem 15.20 were to apply, we would at least predict that such changes after 1000 years will still be of order $\mathcal{O}(\varepsilon)$, so they should, roughly speaking, be noticeable only after a million years.

Now the theorem is not directly applicable, because $\det(\mathrm{D}\omega) = 0$ (in the Kepler problem, orbits of negative energy are always periodic). But there are variants of the theorem that do apply for the problem of celestial mechanics (in FÉJOZ [Fej1], the more extensive question about the applicability of KAM theory, namely permanent stability, is solved).

Indeed, the eccentricity of the orbit of the earth changes under the influence of Jupiter and Saturn in a superposition of periods of about 413 000 and 100 000 years respectively, called the *Milanković cycles*. In a way that is not yet well understood, these cycles are believed to influence the succession of the ice ages.

An early discussion of the stability of the solar system can be found in [Mos1] by MOSER. Numerically this question was studied in high precision by LASKAR, who also included the theory of relativity, the quadrupole moment of the sun, etc., see [Las1]. It was shown that on a time scale of some 10^9 years, collisions among the planets Mercury, Venus, Earth, and Mars can be possible (see [Las2]). ◊

15.22 Exercise (Relativistic Advance of the Perihelion)
Show for the perturbed Kepler problem with the Hamiltonian

$$H_\varepsilon : \widehat{P} \to \mathbb{R} \quad , \quad H_\varepsilon(p,q) = \tfrac{1}{2}\|p\|^2 - \frac{Z}{\|q\|} - \frac{\varepsilon}{\|q\|^3} \tag{15.3.18}$$

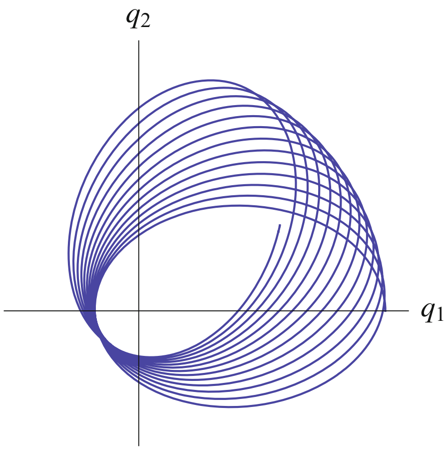
on the phase space $\widehat{P} := T^*\big(\mathbb{R}^2\backslash\{0\}\big)$, that the argument of the Runge-Lenz vector $\hat{A} : \widehat{P} \to \mathbb{R}^2$,

$$\hat{A}(p,q) = \hat{L}(p,q)\begin{pmatrix} p_2 \\ -p_1 \end{pmatrix} - Z\frac{q}{\|q\|}$$

from Exercise 11.22 for a value $E < 0$ of the energy H_0 changes by

$$\varepsilon\,\frac{6\pi Z}{\ell^4} + \mathcal{O}(\varepsilon^2) \tag{15.3.19}$$

within a period of the Kepler ellipse, whereas $\|\hat{A}\|$ (and thus the eccentricity e) remains constant up to order $\mathcal{O}(\varepsilon^2)$. Here $\ell \neq 0$ denotes the value of the angular momentum. ◊



15.23 Remark (Advance of the Perihelion of Mercury)
As perturbations by general relativistic effects appear in the form of a potential $\frac{Z\ell^2}{c^2\|q\|^3}$, the perturbation parameter ε in (15.3.18) is $\varepsilon = \frac{Z\ell^2}{c^2}$. According to (15.3.19) this implies an advance of the perihelion by the approximate angle

$$\frac{6\pi Z^2}{c^2\ell^2} = \frac{6\pi Z}{c^2 a\,(1-e^2)} = 3\pi\frac{r_s}{a\,(1-e^2)}\,,$$

where $r_s = 2Z/c^2$ denotes the Schwarzschild radius of the central mass.

For the sun, $r_s \approx 2.95$ km. The major semiaxis a of Mercury measures about 57 900 000 km, and its eccentricity is $e \approx 0.206$. Thus this effect causes an advance of the perihelion by 2π after about 12 500 000 orbit periods. This effect is less than one percent of the advance in perihelion caused by the other planets. Nevertheless, in 1859, Le Verrier observed this effect, which could not be explained by known causes.

In 1845, Le Verrier had succeeded to calculate the orbit of the planet Neptune from observations of perturbative effects on the planet Uranus, which then led to the discovery of Neptune.

Explaining the advance of Mercury's perihelion by a hypothetical planet, called *Vulcan* by Le Verrier, turned out not to be successful. Rather, this effect became a major confirmation for Einstein's theory of general relativity. ◊

15.24 Literature A reference on the theory of averaging is [SV] by Sanders and Verhulst. A nice collection on a variety of aspects of Hamiltonian dynamics is offered in [MM] by MacKay and Meiss. ◊

15.4 KAM Theory

"On sera frappé de la complexité de cette figure, que je ne cherche même pas à tracer. Rien n'est plus propre à nous donner une idée de la complication du problème des trois corps et en général de tous les problèmes de Dynamique où il n'y a pas d'intégrale uniforme et où les séries de Bohlin sont divergentes."
Henri Poincaré[9]

In the proof of Theorem 15.20 about the first order perturbation theory, we were able to introduce new coordinates $\omega = \omega(I)$ for diophantine frequency vectors such that in these new coordinates, the perturbation terms are of order ε^2 instead of ε.

The question arises whether this transformation can be applied iteratively such as to eliminate the perturbation completely in appropriate coordinates. This would then show the existence of a flow-invariant torus on which the motion is conditionally periodic with frequency ω.

This is the goal of the theory by Kolmogorov, Arnold and Moser established about 1960, briefly called **KAM theory**.

It is related to the classical Newton method for finding zeros of real functions, see Appendix D.

15.25 Remark (Banach's Fixed Point Theorem and Newton Method)
Banach's fixed point method has a speed of convergence that is exponential in the number m of iterations; this means an error bound of $\mathcal{O}(\theta^m)$ (see Theorem D.3).

[9]Translation: "One will be amazed about the complexity of this picture, which I will not even attempt to draw. There is nothing more adequate to give us an idea of the intricacy of the three body problem, and generally all those problems of dynamics in which there are no integrals of motion and where the [perturbation] series by Bohlin diverge."
After: Henri Poincaré: New Methods of Celestial Mechanics, Daniel L. Goroff, Ed., American Institute of Physics. Chapter XXXIII, Paragraph 397
Presumably, Poincaré had something like Figure 15.4.3 on page 428 in mind.

The Newton method converges far more rapidly, with an error of order $\theta^{(2^m)}$ (*quadratic convergence*). However, stronger hypotheses on f are required, namely in particular it has to be differentiable twice.

The following proof of the KAM theorem (Theorem 15.26 on page 415) is due to J. FÉJOZ [Fej2] and uses a fixed point iteration. However, formally, the convergence is a quadratic one, similar to the Newton method.

As shown by Féjoz in [Fej2], one can alternatively use a variation of the Newton method for proving the KAM theorem. ◊

Every proof of this fundamental theorem is complicated, the present one is no exception. After all, KAM theory is "one of the big mathematical achievements" of the 20th century, as commented by Winfried Scharlau in 1992 on the occasion of the award of the Cantor medal to Jürgen Moser.

However, the proof method by Féjoz appears to be comparably transparent.

Afterwards, we will present another result, Theorem 15.33 on page 426. This result shows that it is possible to perform a transformation onto all diophantine tori *simultaneously*. The proof can be found in the article [Poe] by J. PÖSCHEL.

15.4.1 * A Proof of the KAM Theorem

The following proof is based on the one by J. FÉJOZ in [Fej2].

1. The Phase Spaces:

- The real phase space P is of the form $P := \mathbb{R}^n \times \mathbb{T}^n$, with the torus $\mathbb{T}^n = (\mathbb{R}/2\pi\mathbb{Z})^n$. It carries action-angle coordinates (I, φ). The iteration scheme is designed in such a way that in the end, the invariant torus will have action coordinates $I = 0$, i.e., is the zero section $P_0 := \{0\} \times \mathbb{T}^n \subset P$. (Instead of $I = 0$, any other value of the actions could be chosen.)
- The complexified phase space $P_\mathbb{C} \supset P$ is of the form $P_\mathbb{C} := \mathbb{C}^n \times \mathbb{T}^n_\mathbb{C}$, with the complexified torus $\mathbb{T}^n_\mathbb{C} := (\mathbb{C}/2\pi\mathbb{Z})^n$.
- During the iteration, shrinking neighborhoods of the zero section $P_0 \subset P_\mathbb{C}$ will be used. For

$$\mathbb{T}^n_s := \{\varphi \in \mathbb{T}^n_\mathbb{C} \mid \max_{1\le k\le n} |\mathrm{Im}(\varphi_k)| \le s\} \qquad (s > 0),$$

  these neighborhoods are of the form $P_s := \{(I, \varphi) \in P_\mathbb{C} \mid \|(I, \varphi)\| \le s\}$ with

$$\|(I, \varphi)\| := \max_{1\le k\le n} \max\left(|I_k|, |\mathrm{Im}(\varphi_k)|\right). \tag{15.4.1}$$

2. The Hamilton Functions:

For appropriate subsets U, V of vector spaces $\mathbb{C}^k$, we define

$$\mathcal{A}(U,V) := \left\{ g \in C(U,V) \mid g \text{ is real} - \text{analytic in } \mathring{U} \right\} \quad , \quad \mathcal{A}(U) := \mathcal{A}(U,\mathbb{C})\,.$$

Real-analytic means that g is analytic and its restriction to $U \cap \mathbb{R}^k$ is real-valued.

- The Hamilton functions occurring during the iteration will be elements of the $\mathbb{R}$-vector spaces[10]

$$\mathcal{H}_s := \mathcal{A}(P_s) \qquad (s > 0).$$

  As P_s is compact, $\mathcal{A}(P_s)$ with the norm $|H|_s := \sup_{(I,\varphi)\in P_s} |H(I,\varphi)|$ is a Banach space.
- The inductive limit of these Banach spaces is

$$\mathcal{H} \;:=\; \varinjlim \mathcal{H}_s \;=\; \left(\cup_{s>0}\mathcal{H}_s\right)/\!\sim\,,$$

  with the identification $\sim$ of $H_1 \in \mathcal{H}_{s_1}$ and $H_2 \in \mathcal{H}_{s_2}$, whenever $H_1\restriction_{P_s} = H_2\restriction_{P_s}$ for $s := \min\{s_1, s_2\}$. Here the restriction is an *injective* linear mapping since the functions under consideration are analytic. In this sense, one has $\mathcal{H}_t \subset \mathcal{H}_s$ and $|H|_t \geq |H|_s$ for $s \leq t$ and $H \in \mathcal{H}_t$.
- In general, the Hamiltonian vector fields of these functions are not tangential to the torus P_0. For the affine subspaces

$$\mathcal{K}_{s,\omega} := \{H \in \mathcal{H}_s \mid \; \mathrm{D}H|_{P_0} = (\omega, 0)\} \tag{15.4.2}$$

  indexed by the frequency vectors $\omega \in \mathbb{R}^n$, however, H generates the conditionally periodic flow (15.1.2) on P_0.

  Since the only requirement in (15.4.2) is for H to be φ-independent in leading order of I, we have not forfeited the opportunity to transform the (generally non-integrable) Hamilton function $H \in \mathcal{H}_s$ canonically into a function from the space $\mathcal{K}_{s,\omega}$.

3. The Canonical Transformations:

In first order Hamiltonian perturbation theory (Theorem 15.20), small perturbation of an integrable (i.e., only dependent on the actions I) Hamiltonian H_0 were being considered. It was assumed already then that the frequency vector $\omega = \mathrm{D}H_0$ was non-degenerate, i.e., that the matrix $\mathrm{D}\omega = \mathrm{D}^2 H_0$ was regular.

KAM theory consists of the controlled infinite iteration of this theorem. In doing so, while preserving the diophantine condition in every step, this nondegeneracy is of paramount importance, because it permits to adjust the frequency vector by a (symplectic) shift of the action variable.

[10] As the complexified phase space $P_{\mathbb{C}}$ consists of equivalence classes modulo the lattice $(2\pi\mathbb{Z})^n$, we can view H as a function on $\mathbb{C}^n \times \mathbb{C}^n$, $(2\pi\mathbb{Z})^n$-periodic in the second factor.

Roughly speaking, the proof strategy is the following: For nondegenerate $K \in \mathcal{K}_{s,\omega}$ that are integrable on the zero section, and for neighboring $H \in \mathcal{H}$, iterate the mapping

$$(K\,,\;H - K) \longmapsto \big(K + \delta K\,,\;H \circ \exp(-\mathbf{Z}) - (K + \delta K)\big) \tag{15.4.3}$$

to find its fixed point $(K_\infty, 0)$. Here, δK and $\mathbf{Z}$ are the solution of the cohomological equation

$$L_{\mathbf{Z}}\,K + \delta K = H - K\,. \tag{15.4.4}$$

This will not only show that H has an invariant torus of frequency ω, but will also construct it by the application of symplectomorphisms to P_0.

$\exp(-\mathbf{Z})$ is the composition $\Phi_1 \circ \Phi_2 \circ \Phi_3$ of three symplectic mappings:

- The cotangent lift $\Phi_1 := T^*g$ of a diffeomorphism $g : \mathbb{T}^n \to \mathbb{T}^n$ (see page 231) will change the angle coordinates on the torus P_0. In order to achieve uniqueness of $\mathbf{Z}$ when solving (15.4.4), we demand that g arises from a vector field whose average over the torus vanishes.
- The Hamiltonian fiber translation (see Definition 10.36) $\Phi_2 := \text{trans}_{dH}$;
- The symplectic fiber translation $\Phi_3 := \text{trans}_v$ by a constant translation vector $v \in \mathfrak{t}^*$ (with $\mathfrak{t}^* \cong \mathbb{R}^n$ being the dual Lie algebra of $\mathbb{T}^n$, see page 559). For $v \neq 0$, this fiber translation is not Hamiltonian (Figure 15.4.1).

In the illustrations, the effects of these mappings in the case of $n = 1$ degree of freedom is depicted.
The goal is to prove the following theorem about the single diophantine tori.

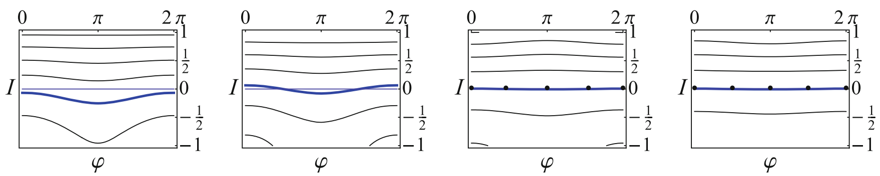


Figure 15.4.1 (1) phase portrait for a Hamilton function of one degree of freedom. The 1-torus with frequency $\sqrt{2}$ is highlighted. (2) After a constant translation of the actions. (3) After a fiber translation. (4) After the cotangent lift.—In parts (3) and (4), we marked equidistant points with respect to time on the torus. In (4), they also have exactly the same distance in angles

15.26 Theorem (KAM) *Let $\omega \in \Omega_{\gamma,\tau}$ (i.e., diophantine, see (15.3.8)).*

- *Let the Hamilton function H_0 be integrable for the frequency vector ω (i.e., $H_0 \in \mathcal{K}_{t,\omega}$, see (15.4.2))*
- *and have a nondegenerate averaged variation of the frequency, i.e., the matrix $\int_{\mathbb{T}^n} \mathrm{D}_1^2 H_0(0,\varphi)\,\mathrm{d}\varphi \in \mathrm{Mat}(n,\mathbb{R})$ is regular.*

Then all Hamilton functions $H \in \mathcal{H}_t$ in a small neighborhood of H_0 also have an invariant torus of frequency ω.

4. The Convergence Conditions:

For $t > 0$, we denote by $\mathcal{A}(\mathbb{T}^n_t)$ the $\mathbb{R}$-vector space of real analytic functions $F : \mathbb{T}^n_t \to \mathbb{C}$. Together with the supremum norm $|\cdot|_t$, the space $\mathcal{A}(\mathbb{T}^n_t)$ is a Banach space. For $H \in \mathcal{A}(\mathbb{T}^n_t)$ with average $\langle H \rangle = 0$, we are looking for a solution S of

$$L_\omega S = H \quad , \quad \langle S \rangle = 0\,, \tag{15.4.5}$$

(called the *cohomological equation*, see Remark 15.17), which involves the Lie derivative with respect to the constant vector field[11] $\omega \in \mathfrak{t} \cong \mathbb{R}^n$. The average $\langle S \rangle$ of S is not determined by the equation $L_\omega S = H$. In Corollary 15.16, the cohomological equation was solved for functions of finite order of differentiability rather than analytic functions.

The Cauchy formula $f(z) = (2\pi i)^{-1} \oint_{\mathcal{C}} \frac{f(w)}{w-z}\,\mathrm{d}w$ for the value of an analytic function in the interior of a closed curve $\mathcal{C}$ in $\mathbb{C}$ implies the formula

$$f'(z) = \frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{f(w)}{(w-z)^2}\,\mathrm{d}w$$

for its derivative. This in turn permits the estimate

$$|f'|_s \le \sigma^{-1} |f|_{s+\sigma} \qquad \big(f \in \mathcal{A}(\mathbb{T}^n_{s+\sigma})\big). \tag{15.4.6}$$

15.27 Lemma (Scalar Cohomological Equation)
If the frequency vector is diophantine with $\omega \in \Omega_{\gamma,\tau}$ and if $H \in \mathcal{A}(\mathbb{T}^n_{s+\sigma_0})$ has average $\langle H \rangle = 0$, then (15.4.5) has in $\mathcal{A}(\mathbb{T}^n_s)$ a unique solution S, and this solution satisfies the norm estimates

$$|S|_s \le \tfrac{C}{\gamma} \sigma^{-\tau-n} |H|_{s+\sigma} \qquad \big(\sigma \in (0, \sigma_0]\big)$$

for a constant $C = C(n,\tau) > 1$.

Proof:

- As in the proof of Corollary 15.16, we use the representations

$$H = \textstyle\sum_{\ell \in \mathbb{Z}^n} H_\ell\, e_\ell \quad , \quad S = \sum_{\ell \in \mathbb{Z}^n} S_\ell\, e_\ell\,, \tag{15.4.7}$$

with the characters $e_\ell(\varphi) := \exp(\imath\, \langle \ell, \varphi \rangle)$ and the Fourier coefficients $H_\ell, S_\ell \in \mathbb{C}$. This representation is applied to H on the domain $\mathbb{T}^n_{s+\sigma}$. From $L_\omega S = \imath \sum_{\ell \in \mathbb{Z}^n} S_\ell \langle \ell, \omega \rangle\, e_\ell$ and $\langle S \rangle = S_0$, one obtains the solution

[11] $\mathfrak{t}$ denotes the Lie algebra of the Lie group $\mathbb{T}^n$.

$$S_\ell := -i\frac{H_\ell}{\langle \ell, \omega\rangle} \qquad \big(\ell \in \mathbb{Z}^n\backslash\{0\}\big) \quad \text{and} \quad S_0 := 0\,. \tag{15.4.8}$$

- Since H and e_ℓ are analytic, the averages $\langle H\, e_\ell\rangle_\theta$ defined by

$$\langle F\rangle_\theta := (2\pi)^{-n}\int_{\mathbb{T}^n+\imath\theta} F(\psi)\,\mathrm{d}\psi \qquad \big(\theta\in\mathbb{R}^n,\ \max_k|\theta_k|\le s+\sigma\big)$$

do not depend on the choice of the vector $i\theta$ by which the real torus $\mathbb{T}^n \subset \mathbb{T}^n_{s+\sigma}$ is shifted. With the optimal choice $\theta_k := -\mathrm{sign}(\ell_k)\,(s+\sigma)\quad (k=1,\ldots,n)$, one obtains a Paley-Wiener estimate (analogous to Lemma 15.14) with a decay of the Fourier coefficients

$$|H_\ell| = |\,\langle H\, e_{-\ell}\rangle_\theta\,| \le \exp\big(-(s+\sigma)|\ell|\big)\ |H|_{s+\sigma} \qquad (\ell\in\mathbb{Z}^n).$$

- With $\omega\in\Omega_{\gamma,\tau}$, hence $|\langle \ell,\omega\rangle|\ge\gamma|\ell|^{-\tau}$, it therefore follows from (15.4.8) that

$$|S_\ell| \le \frac{|\ell|^\tau}{\gamma}\exp\big(-(s+\sigma)|\ell|\big)\ |H|_{s+\sigma} \qquad \big(\ell\in\mathbb{Z}^n\backslash\{0\}\big).$$

With this, we estimate (analogous to Lemma 15.15) the supremum norm $|S|_s$ in $\mathbb{T}^n_s \subset \mathbb{T}^n_{s+\sigma}$. Because of $|\{\ell\in\mathbb{Z}^n \mid |\ell|=k\}| \le 2^n|\{\ell\in\mathbb{N}_0^n \mid |\ell|=k\}| = 2^n\binom{k+n-1}{n-1} \le \frac{2^n}{(n-1)!}(k+n-1)^{n-1}$, we get

$$\begin{aligned}
|S|_s &\le \sum_{\ell\in\mathbb{Z}^n}|S_\ell|\exp\big(s\,|\ell|\big) \le \frac{|H|_{s+\sigma}}{\gamma}\sum_{\ell\in\mathbb{Z}^n\backslash\{0\}}|\ell|^\tau\exp\big(-\sigma|\ell|\big)\\
&\le \tfrac{2^n|H|_{s+\sigma}}{\gamma\,(n-1)!}\ \textstyle\sum_{k\in\mathbb{N}}(k+n-1)^{n-1}\,k^\tau\,e^{-\sigma k}\\
&\le \tfrac{2^n|H|_{s+\sigma}}{\gamma\,(n-1)!}\ \textstyle\sum_{k\in\mathbb{N}}(k+n-1)^{n+\tau-1}\,e^{-\sigma k}\\
&\le |H|_{s+\sigma}\frac{2^n e^{\sigma_0 n}}{\gamma\,(n-1)!}\sum_{m=n+1}^{\infty}(m-1)^{n+\tau-1}e^{-\sigma m}\\
&\le |H|_{s+\sigma}\frac{2^n e^{\sigma_0 n}}{\gamma\,(n-1)!}\int_0^\infty x^{n+\tau-1}e^{-\sigma x}\,\mathrm{d}x\,.
\end{aligned}$$

The integral has the value $\sigma^{-n-\tau}\Gamma(n+\tau)$. So we choose

$$C := \frac{2^n e^{\sigma_0 n}}{(n-1)!}\Gamma(n+\tau). \qquad \square$$

5. The Symplectic Vector Fields:

The above-mentioned symplectomorphisms are represented as solutions of (locally) Hamiltonian differential equations.
This is why we define $\mathbb{R}$-vector spaces parametrized by $s > 0$, whose elements will in turn define the symplectic vector fields.

- The space of real-analytic vector fields whose average over the torus vanishes:

$$\mathcal{X}_s := \left\{X \in \mathcal{A}(\mathbb{T}^n_s, \mathbb{C}^n) \mid \textstyle\int_{\mathbb{T}^n} X \, \mathrm{d}\varphi = 0\right\}.$$

  $\mathcal{X}_s$ acts on the phase space $\mathbb{C}^n \times \mathbb{T}^n_s$ by the Hamiltonian vector fields

$$(I, \varphi) \longmapsto \left(-\mathrm{D}X_\varphi I, X(\varphi)\right).$$

- The space of closed real-analytic 1-forms on the torus:

$$\mathcal{Y}_s := \left\{Y \in \mathcal{A}(\mathbb{T}^n_s, \mathbb{C}^n) \mid dY = 0\right\}.$$

  $\mathcal{Y}_s$ acts on $\mathbb{C}^n \times \mathbb{T}^n_s$ by the symplectic vector fields $(I, \varphi) \mapsto \left(Y(\varphi), 0\right)$.
- The direct sum $\mathcal{Z}_s := \mathcal{X}_s \oplus \mathcal{Y}_s$ of these spaces, with its infinitesimally symplectic action on $\mathbb{C}^n \times \mathbb{T}^n_s$: The mapping $Z = (X, Y) \in \mathcal{Z}_s \subset \mathcal{A}(\mathbb{T}^n_s, \mathbb{C}^{2n})$ that we are looking for acts as a vector field on the phase space $\mathbb{C}^n \times \mathbb{T}^n_s$ in the form

$$\mathbf{Z} : \mathbb{C}^n \times \mathbb{T}^n_s \to \mathbb{C}^{2n} \quad , \quad (I, \varphi) \mapsto \left(Y(\varphi) - \mathrm{D}X_\varphi I, \; X(\varphi)\right). \qquad (15.4.9)$$

$\mathcal{X}_s$, $\mathcal{Y}_s$ and $\mathcal{Z}_s$ are closed subspaces of $\mathcal{A}(\mathbb{T}^n_s, \mathbb{C}^m)$ with the norm

$$\|V\|_s := \max\{|V_k|_s \mid k = 1, \ldots, m\},$$

and as such, they are Banach spaces.

15.28 Lemma (Exponential Map) *For all* $Z = (X, Y) \in \mathcal{Z}_{s+\sigma}$ *with norm* $\|Z\|_{s+\sigma} \leq \frac{1}{3}(\frac{\sigma}{4n})^2$, *the symplectic transformation* $\exp(\mathbf{Z})$ *is real-analytic on* P_s *and translates this phase space domain by no more than*[12]

$$\|\exp(\mathbf{Z}) - \mathbb{1}_{P_s}\|_s \leq \tfrac{6n}{\sigma} \|Z\|_{s+\sigma}. \qquad (15.4.10)$$

Proof:

- The Picard operator for the initial value problem $\dot{x} = \mathbf{Z}(x)$, $x(0) = x_0$ with $x = (I, \varphi) \in P_s$, i.e., in the complex compact phase space, is

$$A : F(\delta) \to \mathcal{A}\left(I_s \times P_s, \mathbb{C}^{2n}\right) \quad , \quad A\psi(t, x) = x + \int_0^t \mathbf{Z}\left(\psi(\tilde{t}, x)\right) \mathrm{d}\tilde{t} \quad (15.4.11)$$

[12] With the notation $t \mapsto \exp(t\mathbf{Z})$ for the flow of the vector field $\mathbf{Z}$.

for $0 < \delta < \sigma$. Hereby,
$F(\delta) := \big\{\psi \in \mathcal{A}\big(I_s \times P_s, \mathbb{C}^{2n}\big) \mid \|\psi(t,\cdot) - \mathrm{Id}_{P_\mathbb{C}}\|_s \le \delta \text{ for all } t \in I_s\big\}$, where I_s is the complexified time interval $I_s := \{t \in \mathbb{C} \mid |\Re(t)| \le 1,\ |\mathrm{Im}(t)| \le s\}$.

- Having specified these domains and ranges, $\mathbf{Z}$ and $\psi(\tilde{t},\cdot)$ can be composed. The line integral is well-defined, because the integrand is analytic.

- On $P_{s+\delta}$, the fiberwise affine vector field $\mathbf{Z}$ is bounded by

$$\|\mathbf{Z}\|_{s+\delta} = \sup\Big\{\big\|\big(Y(\varphi) - \mathrm{D}X_\varphi I,\ X(\varphi)\big)\big\| \ \Big|\ (I,\varphi) \in P_{s+\delta}\Big\} \le \frac{2n}{\sigma-\delta}\|Z\|_{s+\sigma}\,,$$

because $\sigma - \delta < 1$. Choosing now $\delta := \frac{1}{3}\sigma$, one obtains

$$\|\mathbf{Z}\|_{s+2\delta} \le \frac{6n}{\sigma}\|Z\|_{s+\sigma} \le \frac{\sigma}{8n}\,. \tag{15.4.12}$$

Thus, for $(t,x) \in I_s \times P_s$, one has the estimate $\|A\psi\,(t,x) - x\| \le \|\mathbf{Z}\|_{s+2\delta} < \delta$, hence $A\psi \in F(\delta)$. Therefore A can be iterated.

- The Lipschitz constant $\|\mathrm{D}\mathbf{Z}\|_{s+\delta}$ of the Picard operator is obtained from

$$\begin{aligned}\|A\psi_1 - A\psi_0\|_s &\le \sup_{t\in I_s}\Big\|\textstyle\int_0^t\big[\mathbf{Z}\big(\psi_1(\tilde{t},x)\big) - \mathbf{Z}\big(\psi_0(\tilde{t},x)\big)\big]\,\mathrm{d}\tilde{t}\ \Big\|_s\\ &= \sup_{t\in I_s}\Big\|\textstyle\int_0^t\int_0^1 \mathrm{D}\mathbf{Z}\big(\psi_u(\tilde{t},x)\big)\cdot\big(\psi_1(\tilde{t},x) - \psi_0(\tilde{t},x)\big)\,\mathrm{d}u\,\mathrm{d}\tilde{t}\ \Big\|_s\\ &\le \|\mathrm{D}\mathbf{Z}\|_{s+\delta}\|\psi_1 - \psi_0\|_s\,,\end{aligned}$$

where we have written $\psi_u := u\psi_1 + (1-u)\psi_0$ (as in the proof of Lemma 3.14).

- For our choice of δ, the Lipschitz constant is dominated by a Cauchy estimate of the norm of the row sums by (15.4.6):

$$\|\mathrm{D}\mathbf{Z}\|_{s+\delta} \le \tfrac{2n}{\delta}\|\mathbf{Z}\|_{s+2\delta} \le \tfrac{3}{4} < 1\,.$$

- As a contraction on the complete metric space $F(\delta)$, the mapping A has a fixed point Ψ by Banach's fixed point theorem, and $\Psi(t,\cdot) = \exp(t\mathbf{Z}) \quad (t \in [0,1])$.
- Inequality (15.4.12) then also shows (15.4.10). □

Notation: In order to avoid confusion with the scalar product, we will now change the notation for the average $\langle f\rangle$ of f over $\mathbb{T}^n$ into $\overline{f}$, and we will accordingly write $f = \overline{f} + \tilde{f}$. ◊

Now we will address the condition of regular variation of the frequencies in the KAM theorem 15.26, namely

$$\mathrm{D}_1^2\overline{H}_0\restriction_{P_0} \;=\; (2\pi)^{-n}\int_{\mathbb{T}^n}\mathrm{D}_1^2 H_0(0,\varphi)\,\mathrm{d}\varphi \;\in\; \mathrm{GL}(n,\mathbb{R})\,.$$

On a small ball

$$\mathcal{B}_{t,\omega} = \mathcal{B}_{t,\omega}(H_0, \varepsilon_0) := \{K \in \mathcal{K}_{t,\omega} \mid |K - H_0|_t \leq \varepsilon_0\} \tag{15.4.13}$$

around H_0, the variation of the frequencies of K is not degenerate either. We set (in terms of the row sum norm $\|\cdot\|_s$)

$$\begin{aligned} K_{\max} &:= 1 + \max\{|K|_{s_{\max}} \mid K \in \mathcal{B}_{t,\omega}(H_0)\}, \\ C_K &:= K_{\max}\left(1 + \max\left\{\left\|\mathrm{D}_1^2 \overline{K}^{-1}\right\|_{s_{\max}} \mid K \in \mathcal{B}_{t,\omega}(H_0)\right\}\right). \end{aligned} \tag{15.4.14}$$

Just like C_K, in the following lemma, one will also find some constant $\tilde{C}_K$ that depends only on the choice of the ball $\mathcal{B}_{t,\omega}$.
The action on $\mathcal{H}_t$ by the space $\mathcal{Z}_t$, which parametrizes locally Hamiltonian vector fields, is, at H_0, transversal to the integrable subspace $\mathcal{K}_{t,\omega}$:

15.29 Lemma (Solution of the Cohomological Equation)
The cohomological equation $L_{\mathbf{Z}} K + \delta K = \delta H$ for $(K, \delta H) \in \mathcal{B}_{s+\sigma,\omega} \times \mathcal{H}_{s+\sigma}$ has a unique solution $(\delta K, Z) \in \mathcal{K}_{s,0} \times \mathcal{Z}_s$, with the norm

$$\max\big(\|\delta K\|_s\,,\ \|Z\|_s\big) \ \leq\ \tilde{C}_K\ \sigma^{-2(\tau+n+1)}\ |\delta H|_{s+\sigma}\,. \tag{15.4.15}$$

Proof:

- The Taylor expansions of $K \in \mathcal{K}_{s+\sigma,\omega}$ and of the variation $\delta K \in \mathcal{K}_{s+\sigma,0}$ of K in the actions I are

$$\begin{aligned} K(I,\varphi) &= K^{(0)} + \langle \omega, I\rangle + \big\langle I,\ K^{(2)}(\varphi)\, I\big\rangle + K^{(3)}(I,\varphi)\,, \\ \delta K(I,\varphi) &= \delta K^{(0)} \qquad\qquad\quad + \ \delta K^{(2)}(I,\varphi)\,, \end{aligned}$$

with $\langle a, b\rangle = \sum_{k=1}^n a_k b_k$ for $a, b \in \mathbb{C}^n$, and $K^{(2)}(\varphi) \in \mathrm{Sym}(n, \mathbb{C})$; one also has $K^{(3)}(I,\varphi) = \mathcal{O}(\|I\|^3)$ and $\delta K^{(2)}(I,\varphi) = \mathcal{O}(\|I\|^2)$. Therefore

$$\mathrm{D}K(I,\varphi) = \big(\omega + 2K^{(2)}(\varphi)\, I,\ 0\big) + \mathcal{O}(\|I\|^2)\,. \tag{15.4.16}$$

- With (15.4.9) and (15.4.16), the first term $L_{\mathbf{Z}} K = \mathrm{D}K(\mathbf{Z})$ of the cohomological equation has the Taylor expansion

$$L_{\mathbf{Z}} K(I,\varphi) = \big\langle \omega + 2K^{(2)}(\varphi)\, I, Y(\varphi) - \mathrm{D}X_\varphi I\big\rangle + \mathcal{O}(\|I\|^2)\,.$$

- In contrast, the leading term of the δH expansion may already depend on φ:

$$\delta H(I,\varphi) = \delta H^{(0)}(\varphi) + \big\langle \delta H^{(1)}(\varphi), I\big\rangle + \delta H^{(2)}(I,\varphi)$$

with $\delta H^{(2)}(I,\varphi) = \mathcal{O}(\|I\|^2)$.

- Comparing coefficients in the I-expanded cohomological equation yields:

$$\langle \omega, Y(\varphi)\rangle + \delta K^{(0)} = \delta H^{(0)}(\varphi) \quad , \quad 2K^{(2)}(\varphi)Y(\varphi) - \mathrm{D}X_{\varphi}^{\top}\,\omega = \delta H^{(1)}(\varphi)\,,$$

and $\delta K^{(2)}(I,\varphi) = \delta H^{(2)}(I,\varphi) + \big\langle 2K^{(2)}(\varphi)\,I, \mathrm{D}X_{\varphi} I\big\rangle - \mathrm{D}K^{(3)}(\mathbf{Z})(I,\varphi)$.
- This linear system of equations has the unique solution

$$\tilde{Y} := \mathrm{D}\,L_{\omega}^{-1}\big(\delta\tilde{H}^{(0)}\big) \quad , \quad \overline{Y} := \tfrac{1}{2}\big(\overline{K}^{(2)}\big)^{-1}\big(\delta\overline{H}^{(1)} - 2\overline{K^{(2)}\tilde{Y}}\big)\,, \tag{15.4.17}$$

$$X := L_{\omega}^{-1}\big(2K^{(2)}Y - \delta H^{(1)}\big) \quad , \quad \delta K^{(0)} := \delta\overline{H}^{(0)} - \big\langle \omega, \overline{Y}\big\rangle \tag{15.4.18}$$

and, using these quantities, the $\delta K^{(2)}$ already mentioned above.

- In order to control the norms of these quantities, they will be estimated in a sequence of shrinking domains, in their order of occurrence. With $\omega \in \Omega_{\gamma,\tau}$ and the constant $C > 1$ from Lemma 15.27, one has therefore

$$\|\tilde{Y}\|_{s+\frac{2}{3}\sigma} \le \tfrac{C}{\gamma}(\sigma/3)^{-\tau-n-1}|\delta\tilde{H}|_{s+\sigma} \qquad \big(\sigma \in (0,\sigma_0]\big).$$

Plugging this into formula (15.4.17) for $\overline{Y}$, one gets (with $C_K \ge 1$ from (15.4.14)):

$$\|\overline{Y}\|_{s+\frac{2}{3}\sigma} \le C_K\big(1 + \tfrac{C}{\gamma}(\sigma/3)^{-\tau-n-1}\big)\,|\delta H|_{s+\sigma} \qquad \big(\sigma \in (0,\sigma_0]\big).$$

Therefore,

$$\|Y\|_{s+\frac{2}{3}\sigma} \le \|\overline{Y}\|_{s+\frac{2}{3}\sigma} + \|\tilde{Y}\|_{s+\frac{2}{3}\sigma} \le 2C_K\big(1 + \tfrac{C}{\gamma}(\sigma/3)^{-\tau-n-1}\big)|\delta H|_{s+\sigma}\,. \tag{15.4.19}$$

Next we estimate the vector field X from (15.4.18) componentwise, by means of Lemma 15.27, and with $K_{\max}$ from (15.4.14), (15.4.19) and $\sigma \le 1$, we obtain:

$$\begin{aligned}\|X\|_{s+\frac{1}{3}\sigma} &\le 2\tfrac{C}{\gamma}(\sigma/3)^{-\tau-n}\,\big(K_{\max}\|Y\|_{s+\frac{2}{3}\sigma} + \|\delta H\|_{s+\sigma}\big)\\ &\le 2\tfrac{C}{\gamma}\,(\sigma/3)^{-\tau-n}\big(2C_K K_{\max}\big(1 + \tfrac{C}{\gamma}(\sigma/3)^{-\tau-n-1}\big) + 1\big)\,\|\delta H\|_{s+\sigma}\\ &\le 8\Big[C_K\big(1 + \tfrac{C}{\gamma}(\sigma/3)^{-\tau-n-1/2}\big)\Big]^2|\delta H|_{s+\sigma}\,.\end{aligned} \tag{15.4.20}$$

Altogether, (15.4.19) and (15.4.20) prove the part of inequality (15.4.15) that concerns $\|Z\|_s \le \|Z\|_{s+\sigma/3}$. Next we deal with the part for $\|\delta K\|_s$:
- The constant function $\delta K^{(0)}$ from (15.4.18) is bounded by

$$|\delta K^{(0)}|_s \le |\delta H|_s + \|\omega\|\,\|Y\|_s \le \Big[1 + 2\|\omega\|C_K\big(1 + \tfrac{C}{\gamma}(\sigma/3)^{-\tau-n-1}\big)\Big]|\delta H|_{s+\sigma}\,.$$

Finally, on the compact phase space domain P_s, we can estimate

$$|\delta K^{(2)}|_s \le |\delta H^{(2)}|_s + |K|_s \, \|\mathrm{D}X\|_s + \|\mathrm{D}K\|_s \|\mathbf{Z}\|_s$$
$$\le |\delta H|_s + \tfrac{3n}{\sigma}|K|_s \, \|X\|_{s+\frac{1}{3}\sigma} + \tfrac{n}{\sigma}\|K\|_{s+\sigma}\|\mathbf{Z}\|_s \, . \qquad (15.4.21)$$

Next to the estimate (15.4.20) for X, we also plug the inequality $\|\mathbf{Z}\|_s \le \frac{6n}{\sigma}\|Z\|_{s+\sigma/3}$, which is the analog of (15.4.12), into (15.4.21). □

Next we estimate how much the Hamiltonian $K + \delta H$ still differs from integrability after it has been transformed by the symplectomorphism $\exp(-\mathbf{Z})$.

15.30 Lemma (Quadratic Convergence)
The solutions $(\delta K, Z) \in \mathcal{K}_{s,0} \times \mathcal{Z}_s$ *of the cohomological equation* $L_{\mathbf{Z}}\, K + \delta K = \delta H$ *for* $(K, \delta H) \in \mathcal{B}_{s+\sigma,\omega} \times \mathcal{H}_{s+\sigma}$ *with* $|\delta H|_{s+\sigma} \le g\sigma^{2\tilde{\tau}}$, $\tilde{\tau} := 2(\tau + n + 2)$ *satisfy the following for an appropriate* g*:*

$$\exp(\mathbf{Z}) \text{ is real-analytic on } P_s \quad , \quad \|\exp(\mathbf{Z}) - \mathrm{Id}_{P_{\mathbb{C}}}\|_s \le \sigma \, . \qquad (15.4.22)$$

The error in the iteration is then

$$|(K + \delta H) \circ \exp(-\mathbf{Z}) - (K + \delta K)|_s \;\le\; c\; (\sigma^{-\tilde{\tau}})^2 \; |\delta H|^2_{s+\sigma} \, . \qquad (15.4.23)$$

Proof:

- Lemmas 15.28 and 15.29 show the correctness of (15.4.22). Indeed, with the given bound for $|\delta H|_{s+\sigma}$, Lemma 15.29 asserts that

$$\max\bigl(\|\delta K\|_{s+\sigma/2} \, , \; \|Z\|_{s+\sigma/2}\bigr) \;\le\; \tilde{C}_K \; (\sigma/2)^{-2(\tau+n+1)} \; |\delta H|_{s+\sigma} \le \tfrac{1}{3}\bigl(\tfrac{\sigma/2}{4n}\bigr)^2 \, ,$$

for $g := \frac{2^{-2(\tau+n+2)}}{3\tilde{C}_K (4n)^2}$. Using (15.4.10), one concludes

$$\|\exp(\mathbf{Z}) - \mathbb{1}_{P_s}\|_s \;\le\; \tfrac{6n}{\sigma/2}\|Z\|_{s+\sigma/2} \;\le\; \tfrac{\sigma}{16n} \, .$$

- For $H := K + \delta H$, the fundamental theorem of calculus yields

$$H \circ \exp(-\mathbf{Z}) = H + \int_0^1 \tfrac{\mathrm{d}}{\mathrm{d}t} H \circ \exp(-t\mathbf{Z})\, \mathrm{d}t$$
$$= H - \int_0^1 \mathrm{D}H_{\exp(-t\mathbf{Z})}(\mathbf{Z})\, \mathrm{d}t = H - \mathrm{D}H(\mathbf{Z}) + R \, ,$$

with the remainder term $R = \int_0^1 \int_0^t \frac{\mathrm{d}}{\mathrm{d}s}\mathrm{D}H \circ \exp(-s\mathbf{Z})(\mathbf{Z})\, \mathrm{d}s\, \mathrm{d}t =$

$$\int_0^1 \int_0^t \mathrm{D}^2 H_{\exp(-s\mathbf{Z})}(\mathbf{Z}, \mathbf{Z})\, \mathrm{d}s\, \mathrm{d}t = \int_0^1 (1-t)\mathrm{D}^2 H_{\exp(-t\mathbf{Z})}(\mathbf{Z}, \mathbf{Z})\, \mathrm{d}t \, . \qquad (15.4.24)$$

The left hand side $|H \circ \exp(-\mathbf{Z}) - (K + \delta K)|_s$ of (15.4.23) is therefore equal to

$$|\delta H - \mathrm{D}H(\mathbf{Z}) + R - \delta K|_s = |R - L_{\mathbf{Z}}\, \delta H|_s \; ; \qquad (15.4.25)$$

this latter identity follows from the cohomological equation $L_{\mathbf{Z}}\, K + \delta K = \delta H$.

- In order to prove (15.4.23), we estimate the right hand side of (15.4.25). From (15.4.12) and (15.4.15), one has

$$\|\mathbf{Z}\|_s \le \tfrac{2n}{\sigma/2}\|Z\|_{s+\sigma/2} \le 2n\tilde{C}_K\,(\sigma/2)^{-2(\tau+n)-3}\,|\delta H|_{s+\sigma}\,. \tag{15.4.26}$$

For $H \in \mathcal{B}_{t,\omega}$ (the ball about H_0 defined in (15.4.13)), the term D^2H in (15.4.24) is estimated by $2n\sigma^{-2}K_{\max}$ with the constant $K_{\max}$ from (15.4.14). The contribution from R to (15.4.25) is therefore of the form given in (15.4.23).

By (15.4.26), the term $|L_{\mathbf{Z}}\,\delta H|_s$ in (15.4.25) is controlled by a constant times $\sigma^{-\tilde{\tau}}\,|\delta H|^2_{s+\sigma}$, hence by a smaller power of $1/\sigma$. □

Féjoz finds the integrable Hamiltonian of the KAM torus as a fixed point $(K_\infty, 0)$ of the mapping (15.4.3). One may therefore believe that it is Banach's fixed point theorem (see D.3) that guarantees the existence of the fixed point.

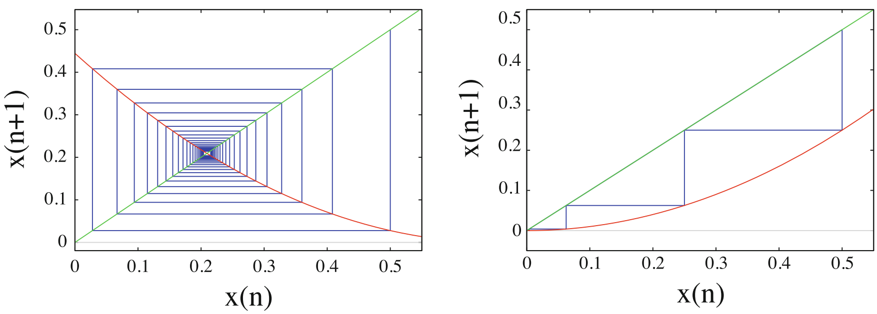


Figure 15.4.2 Iteration of the quadratic mapping F_c for $c = 2/3$ (left) and $c = 0$ (right). The iteration is represented as a *cobweb diagram*, which is a sequence of segments whose endpoints are on the graph and on the diagonal

This is however not possible, because Lemma 15.29 cannot guarantee for the mapping to be a contraction. This is because the sum $\sum_{j=1}^{\infty}\sigma_j$ of the losses in analyticity σ_j in the jth iteration has to be finite, so the σ_j will converge to zero. On the other hand, this will make the factor $\sigma_j^{-2(\tau+n+1)}$ of the upper bound in (15.4.15) diverge.

The situation is salvaged by a particular property of the iterated map. This map becomes more and more contractive the closer one approaches its fixed point. This can be seen from the upper bound in (15.4.23), which is quadratic in δH.

15.31 Remark (Quadratic Convergence to a Fixed Point)
A vastly simplified model is given by the iteration of the quadratic mappings

$$F_c : \mathbb{R} \to \mathbb{R} \quad , \quad x \mapsto (x-c)^2 \qquad \big(c \in [-1/4, \infty)\big).$$

Here the fixed point $x_c := c + 1/2 - \sqrt{c + 1/4} \geq 0$ of F_c is stable for values of the parameter $c \in (-1/4, 3/4)$, hence F_c is contracting in a neighborhood of x_c. However, for $c = x_c = 0$, the derivative $F_c'(x_c)$ equals 0, and the convergence to x_c is quadratic as in the Newton method (Figure 15.4.2, right). $\Diamond$

- However, in contrast to the F_0 from the remark, the derivative of the mapping (15.4.3) at the fixed points is not zero, because *all* integrable Hamiltonians K yield fixed points $(K, 0)$.
- Moreover, the domain of the used mapping shrinks with every iteration.

In the following fixed point theorem, the situation is stated formally.

There, the two function spaces underlying the mapping (15.4.3) are denoted generally as $\mathcal{H}$ and $\mathcal{K}$. More precisely, the quality of analyticity that deteriorates under the iteration is abstractly encoded in a family $(\mathcal{H}_s)_{s\in(0,1]}$ of Banach spaces $\big(\mathcal{H}_s, |\cdot|_s^{(\mathcal{H})}\big)$, with $\mathcal{H}_t \subseteq \mathcal{H}_s$ and $|H|_s^{(\mathcal{H})} \leq |H|_t^{(\mathcal{H})}$ for $s < t$ and $H \in \mathcal{H}_t$. Analogous statements apply for $\big(\mathcal{K}_s, |\cdot|_s^{(\mathcal{H})}\big)_{s\in(0,1]}$ and the cartesian products $\mathcal{M}_s := \mathcal{K}_s \times \mathcal{H}_s$, with the norms $|(K, H)|_s := \max(|K|_s^{(\mathcal{K})}, |H|_s^{(\mathcal{H})})$ (we abbreviate here by writing H instead of δH).

Let $0 < s < s + \sigma \leq 1$ and $h, k, \tau > 0$ be the parameters for the balls

$$B_{s,\sigma}(h, k, \tau) := \{(K, H) \in \mathcal{M}_{s+\sigma} \mid |K|_{s+\sigma}^{(\mathcal{K})} \leq k,\ |H|_{s+\sigma}^{(\mathcal{H})} \leq h\sigma^{\tau}\}. \tag{15.4.27}$$

Assume we are given a family of mappings

$$F_{s,\sigma} : B_{s,\sigma}(h, k, \tau) \longrightarrow \mathcal{M}_s\,,$$

which are compatible with restrictions to subspaces, i.e.,

$$F_{s_1,\sigma_1}\restriction_{\mathcal{M}_{s_2+\sigma_2}} = F_{s_2,\sigma_2} \quad \text{for} \quad 0 < s_1 < s_2 \quad \text{and} \quad 0 < \sigma_1 < \sigma_2\,.$$

The deterioration in quality (which in the application is the shrinking of the domain of analyticity) under $F_{s,\sigma}$ is measured by the parameter $\sigma \in (0, 1 - s]$.

15.32 Theorem (Féjoz [Fej2]; Existence of a Fixed Point)
Assume that, with appropriate $c, h_0, k_0, \hat{\tau} > 0$ and $\tau \geq \hat{\tau}$, the images $(\hat{K}, \hat{H}) := F_{s,\sigma}(K, H)$ of $(K, H) \in B_{s,\sigma}(h_0, k_0, \tau)$ satisfy

$$|\hat{K} - K|_s^{(\mathcal{K})} \leq c\,\sigma^{-\hat{\tau}/2}\,|H|_{s+\sigma}^{(\mathcal{H})} \quad \textit{and} \quad |\hat{H}|_s^{(\mathcal{H})}) \leq c\,\sigma^{-\hat{\tau}}\big(|H|_{s+\sigma}^{(\mathcal{H})}\big)^2 \tag{15.4.28}$$

for all $0 < s < s + \sigma \leq 1$.
Then, for $\sigma_j := 2^{-j}\sigma$, $s_0 := s + \sigma$, for the decreasing sequence $(s_j)_{j\in\mathbb{N}_0}$ defined as $s_j := s_{j-1} - \sigma_j$ (hence with limit s), and for $k := k_0/2$, there exists $h \in (0, h_0)$, such that:

- *The mappings* $\tilde F^{(j)} : B_{s_1,\sigma_1}(h,k,\tau) \to B_{s_{j+1},\sigma_{j+1}}(h,k_0,\tau)$ *with*

$$\tilde F^{(1)} := F_{s_1,\sigma_1} \quad \text{and} \quad \tilde F^{(j)} = F_{s_j,\sigma_j} \circ \tilde F^{(j-1)}$$

have the claimed ranges (i.e., are defined for all $j \in \mathbb{N}$*).*

- *For all* $(K,H) \in B_{s,\sigma}(h,k,\tau)$*, the limit*

$$(K_\infty, 0) \;:=\; \lim_{j\to\infty} \tilde F^{(j)}(K,H) \;\in\; B_{s,0}(h,k_0,\tau)$$

exists.

Proof:

- We have $s_j = s + \sigma - \sum_{l=1}^{j} \sigma_l = s + 2^{-j}\sigma$. Let

$$h := \min\bigl(2^{\hat\tau - 2\tau}/c,\; k\,(1 - 2^{-\tau/2})\bigr)\,.$$

As long as the mappings with the given domains $\tilde F^{(j)}$ are defined with the given ranges, we will write (K_j, H_j) for $\tilde F^{(j-1)}(K,H)$.
For $(K,H) \in B_{s_j,\sigma_j}(h,k,\tau)$, thus $|H_j|^{(\mathcal{H})}_{s_{j-1}} \le h\sigma_j^\tau$, one has by (15.4.28):

$$|H_{j+1}|^{(\mathcal{H})}_{s_j} \;\le\; c\sigma_j^{-\hat\tau} h^2 \sigma_j^{2\tau} \;=\; c\, 2^{2\tau-\hat\tau} h^2 \sigma_{j+1}^{2\tau-\hat\tau} \;\le\; h\sigma_{j+1}^{\tau}\,. \tag{15.4.29}$$

Thus, for this choice of h, all $\tilde F^{(j)}$ have, as far as the norms of the H_j are concerned, the claimed ranges. If the same is true for the K_j, one also has $\lim_{j\to\infty} |H_j|^{(\mathcal{H})}_s = 0$.
- Indeed, $(K_j)_{j\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{K}_s$ with $|K_j|^{(\mathcal{K})}_s \le k_0$, because $\mathcal{K}_{s_j} \subset \mathcal{K}_s$, and because for all $j \ge i \in \mathbb{N}$, one gets from (15.4.29) that $|K_j - K_i|^{(\mathcal{K})}_s \le$

$$\textstyle\sum_{\ell=i}^{\infty} |K_{\ell+1} - K_\ell|^{(\mathcal{K})}_s \;\le\; \sum_{\ell=i}^{\infty} |K_{\ell+1} - K_\ell|^{(\mathcal{K})}_{s_{\ell+1}} \;\le\; h\sigma^\tau \frac{2^{-i\tau/2}}{1-2^{-\tau/2}} \;\le\; k\,. \qquad \square$$

Proof of Theorem 15.26 (KAM) from page 415:
By Lemma 15.29 and (15.4.22), the family of mappings

$$\begin{aligned} F_{s,\sigma} : \mathcal{B}_{s+\sigma,\omega} \times \mathcal{H}_{s+\sigma} &\longrightarrow \mathcal{K}_{s,\omega} \times \mathcal{H}_s \qquad\qquad (0 < s < s+\sigma \le 1) \\ (K\,,\; H - K) &\longmapsto \bigl(K + \delta K\,,\; H \circ \exp(-\mathbf{Z}) - (K + \delta K)\bigr) \end{aligned}$$

from (15.4.3) has the range as claimed. It needs to be shown that, after translation by H_0, this family satisfies the hypothesis of the fixed point theorem 15.32.
Then, if it gets iterated starting with $(H_0, H - H_0)$ and for $s := \sigma := \frac{1}{2}\min(t,1)$, one finds a fixed point $(K_\infty, 0) \in \mathcal{B}_{s,\omega} \times \mathcal{H}_s$.

Because of Lemma 15.29 and estimate (15.4.23) in Lemma 15.30, condition (15.4.28) is satisfied for $\hat\tau := 2\tilde\tau$ and small radii ε_0 of the ball $\mathcal{B}_{s+\sigma,\omega}$ in (15.4.13).$\square$
Similarly, as shown by J. FÉJOZ in [Fej2], the concatenated symplectomorphisms of the KAM iteration converge to a limit symplectomorphism.

15.4.2 Measure of the KAM Tori

We have just developed the KAM theory for the *single* tori with a diophantine frequency vector. In contrast, PÖSCHEL in [Poe] constructed a global canonical transformation that *simultaneously* yields action-angle coordinates for many KAM tori. Moreover, it permits a certain control of the possibly chaotic dynamics outside the KAM tori. Again, let

$$H_\varepsilon(I,\varphi) := H_0(I) + \varepsilon H_1(I,\varphi) \tag{15.4.30}$$

on the phase space $G \times \mathbb{T}^n$, with a domain $G \subset \mathbb{R}^n$ and perturbation parameter $|\varepsilon| < \varepsilon_0$. We assume that the integrable Hamiltonian H_0 is real-analytic. Furthermore, we again assume the independent variation of the frequencies $\omega := \mathrm{D}H_0 : G \to \mathbb{R}^n$,

$$\det(\mathrm{D}\omega)(I) \neq 0 \qquad (I \in G), \tag{15.4.31}$$

and that ω is a diffeomorphism onto its image

$$\Omega := \omega(G)\,.$$

Due to this property, we can use the mapping

$$\Psi : \Omega \times \mathbb{T}^n \to G \times \mathbb{T}^n \quad , \quad (\hat{\omega},\varphi) \mapsto \big(\omega^{-1}(\hat{\omega}),\varphi\big)$$

to view the Hamiltonian H_ε as a function

$$\mathcal{H}_\varepsilon := H_\varepsilon \circ \Psi : \ \Omega \times \mathbb{T}^n \longrightarrow \mathbb{R}$$

and pull back the symplectic 2-form with Ψ^* to this new phase space.

The advantage of this change of coordinates is that during the iteration of the canonical transformation, one can immediately read off from the new actions $\hat{\omega}$ whether they belong to the diophantine set $\Omega_{\gamma,\tau}$ (see (15.3.8)).

We set $\tau := n$ and simply denote the diophantine subset of the set Ω of frequencies by

$$\Omega_\gamma := \Omega_{\gamma,\tau} \cap \Omega\,.$$

The perturbation term H_1 is assumed to be smooth[13] on $G \times \mathbb{T}^n$, i.e., $\mathcal{H}_1 = H_1 \circ \Psi \in C^\infty(\Omega \times \mathbb{T}^n)$.

15.33 Theorem (Kolmogorov, Arnol'd and Moser; KAM)
Under the above hypotheses, there exists $\varepsilon_0 > 0$ such that for $|\varepsilon| < \varepsilon_0$, there exists a diffeomorphism $\mathcal{T}_\varepsilon$ on the phase space $\Omega \times \mathbb{T}^n$ which, on the subset

$$\Omega_{\sqrt{\varepsilon}} \times \mathbb{T}^n \subset \Omega \times \mathbb{T}^n\,,$$

[13] Actually, in [Poe], PÖSCHEL only needs $3n$-fold continuous differentiability.

transforms the Hamiltonian differential equations into the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\omega}(t) = 0 \quad , \quad \frac{\mathrm{d}}{\mathrm{d}t}\varphi(t) = \hat{\omega}(t) \; . \tag{15.4.32}$$

The Lebesgue measure of this subset is large for small perturbations:

$$\lambda^n\left(\Omega_{\sqrt{\varepsilon}}\right) = \lambda^n(\Omega)\cdot\left(1 - \mathcal{O}(\sqrt{\varepsilon})\right)$$

(provided $\overline{\Omega}$ is a manifold with boundary).

15.34 Remarks

1. The equations (15.4.32) are integrable and have the solutions

$$\hat{\omega}(t) = \hat{\omega}(0) \quad , \quad \varphi(t) = \varphi(0) + \hat{\omega}(0)t \ (\mathrm{mod}\ 2\pi) \qquad (t \in \mathbb{R}).$$

2. This version of the theorem tells nothing about the 'resonant tori', i.e., the complement of $\Omega_{\sqrt{\varepsilon}} \times \mathbb{T}^n$. There the tori could either survive (namely if H_ε is integrable) or they could dissolve under the perturbation, see Figure 15.4.3 and the cover picture of this book. ◊

15.35 Example (Motion in a Periodic Potential)
As an applied example, we resume the motion of a particle in an $\mathcal{L}$-periodic potential V, which was discussed in Chapter 11.2. Hereby the regular lattice $\mathcal{L} = \mathrm{span}_{\mathbb{Z}}(\ell_1, \ldots, \ell_d)$ is spanned by a basis $\ell_1, \ldots, \ell_d$ of $\mathbb{R}^d$. We additionally assume smoothness of the potential, i.e., $V \in C^\infty(\mathbb{R}^d, \mathbb{R})$.

The Hamiltonian $H : P \to \mathbb{R}$, $H(p,q) = \frac{1}{2}\|p\|^2 + V(q)$ on the phase space $P = \mathbb{R}^d_p \times \mathbb{R}^d_q$ generates the Hamiltonian differential equation

$$\dot{p} = -\nabla V(q) \quad , \quad \dot{q} = p \, .$$

In order to agree with the above wording of the KAM theorem, we use, differently than in Chapter 11.2, the standard torus $\mathbb{T}^d = \mathbb{R}^d/(2\pi\mathbb{Z})^d$ rather than the period torus $\mathbb{R}^d/\mathcal{L}$. Thus the phase space is the cotangent space of the standard torus. $\hat{\mathcal{P}} := T^*\mathbb{T}^d \cong \mathbb{R}^d \times \mathbb{T}^d$. Using the matrix $L := (\ell_1, \ldots, \ell_d)/(2\pi) \in \mathrm{Mat}(d, \mathbb{R})$ of the basis vectors for a change of coordinates, V becomes 2π-periodic, i.e., it can be written as a function

$$\hat{V} \in C^\infty(\mathbb{T}^d, \mathbb{R}) \quad , \quad \hat{V}(\varphi) := V(L\varphi) \, .$$

In order to apply KAM theory to H with energies lying in the interval $[(1-\delta)E, (1+\delta)E]$ around $E > 0$, we consider the Hamiltonian

$$\hat{H}_\varepsilon : \hat{\mathcal{P}} \to \mathbb{R} \quad , \quad \hat{H}_\varepsilon(I, \varphi) := \tfrac{1}{2}\langle I, MI\rangle + \varepsilon\hat{V}(\varphi) \, ,$$

with $M := (L^\top L)^{-1}$. This function matches the form (15.4.30). For the diffeomorphism

$$\mathcal{M}_E : \mathcal{P} \to \hat{\mathcal{P}} \quad , \quad (p,q) \mapsto (I,\varphi) := \Big(L^\top p/\sqrt{E}\,,\ L^{-1}q\Big)\,,$$

it follows that

$$E \cdot \hat{H}_{1/E} \circ \mathcal{M}_E = H\,,$$

and the Hamiltonian flow $\hat{\Phi}_\varepsilon$ generated by $\hat{H}_\varepsilon$ (with respect to the canonical symplectic structure ω_0 on $\hat{\mathcal{P}}$) is conjugate to the original flow, up to a change in the time scale:

$$\hat{\Phi}_{1/E}^{\sqrt{E}t} \circ \mathcal{M}_E = \mathcal{M}_E \circ \Phi^t \qquad (t \in \mathbb{R}).$$

For perturbation parameter $\varepsilon = 0$, the flow $\hat{\Phi}_\varepsilon^t$ is integrable. In this case,

$$\hat{\Phi}_0^t(I_0,\varphi_0) = \big(I_0, \varphi_0 + \omega(I_0)t\big) \qquad \big(t \in \mathbb{R}, (I_0,\varphi_0) \in \hat{\mathcal{P}}\big)\,,$$

with the frequency vector $\omega(I) := \mathrm{D}\hat{H}_0(I) = MI$. The frequency vector ω does satisfy the condition (15.4.31) of independent variation, because the matrix $\mathrm{D}\omega(I) = M$ has rank d. ◊

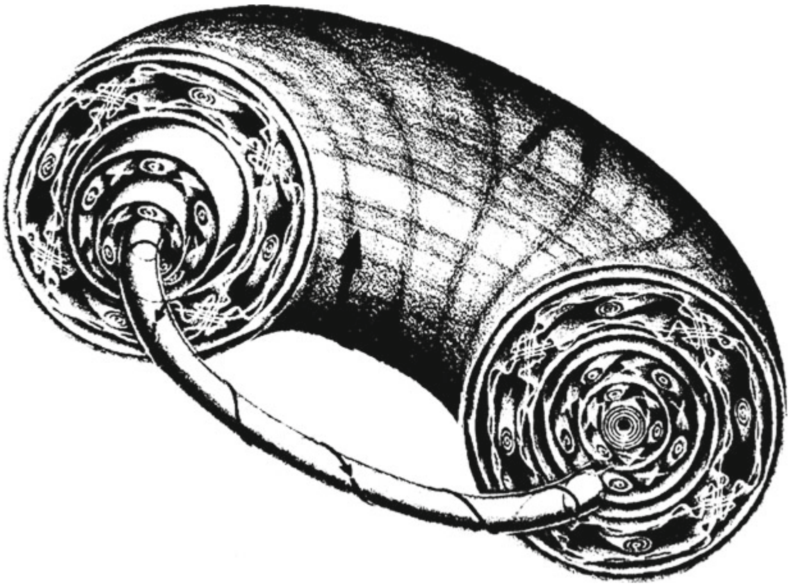


Figure 15.4.3 Phase space portrait of a Hamiltonian system with two degrees of freedom, with nested KAM tori. (Picture: Ralph Abraham and Jerrold E. Marsden, from: Foundations of Mechanics. Addison-Wesley Publishing Company. lnc. 1982, second edition, fourth printing, Figure 8.3–3 [AM], c American Mathematical Society 2008.)

15.36 Remark (Resonances) As will be elaborated on in Chapter 17.2 on the Poincaré-Birkhoff theorem, new stable and unstable periodic orbits frequently occur between the KAM tori (see Figure 15.4.3). They correspond to resonances between frequencies of the perturbed integrable system.

In this spirit, one frequently finds the rotation periods of certain planets and moons in resonance with their orbit periods. So whereas the moon of the earth has a bound rotation, i.e., faces us with the same side all the time, Mercury rotates about its axis exactly three times during two orbit periods.[14] ♢

If one is interested in the portion of KAM tori on one energy surface, the hypotheses of the theorem need to be adjusted. This appears to be impossible, because now, on an $(n-1)$-dimensional energy surface, not all n frequencies can be varied independently. But as can be seen from the diophantine condition (15.3.8), the main issue is that their *ratios* can be varied independently. So one passes from the space $\mathbb{R}^n$ of frequencies to the projective space $\mathbb{RP}(n-1)$.

15.37 Theorem (Isoenergetic Nondegeneracy) *For the integrable Hamiltonian H_0, let $E := H_0(I_0)$ be a regular value of H_0. It is if and only if*

$$\det\begin{pmatrix} D\omega(I_0) & \omega(I_0) \\ \omega(I_0)^\top & 0 \end{pmatrix} \neq 0 \tag{15.4.33}$$

for $\omega := \nabla H_0 : G \to \mathbb{R}^n$ that the mapping

$$U \to \mathbb{RP}(n-1) \quad , \quad I \mapsto \mathrm{span}\big(\omega(I)\big) \tag{15.4.34}$$

is a diffeomorphism onto its image for an appropriate neighborhood $U \subseteq H_0^{-1}(E)$ of I_0.

Proof:

- As E was assumed to be a regular value of H_0, the frequency vector $\omega : G \to \mathbb{R}^n$ will not take on a value 0 on the $(n-1)$-dimensional manifold $M_E := H_0^{-1}(E) \subset G$. Therefore, for all $I \in M_E$, the space $\mathrm{span}\big(\omega(I)\big) \subseteq \mathbb{R}^n$ is a 1-dimensional subspace, and
$$\Phi : M_E \to \mathbb{RP}(n-1) \quad , \quad I \mapsto \mathrm{span}\big(\omega(I)\big)$$
indeed maps into the projective space $\mathbb{RP}(n-1)$.
- The vectors $v \in \mathbb{R}^n$ that are orthogonal to $\omega(I_0)$ form the tangent space $T_{I_0}M_E$ of M_E. It is exactly when there exists such a vector $v_0 \neq 0$ whose image under $D\omega(I_0)$ lies in $\mathrm{span}\big(\omega(I_0)\big)$ that the linear mapping
$$T_{I_0}M_E \to T_{\Phi(I_0)}\mathbb{RP}(n-1) \quad , \quad v \mapsto D\Phi(I_0)\, v$$
is not surjective.

[14] In [CL], Correia and Laskar analyze capture into such resonances by tidal interactions.

- But in this case, the condition (15.4.33) is violated. Indeed, for $\mathrm{D}\omega(I_0)\,v_0 = \lambda\,\omega(I_0)$, the vector $\binom{v_0}{-\lambda}$ will then be in the kernel of the matrix $\begin{pmatrix} \mathrm{D}\omega(I_0) & \omega(I_0) \\ \omega(I_0)^\top & 0 \end{pmatrix}$. Conversely, a vector $\binom{v}{\mu}$ can only be in the kernel of this matrix if $\langle v, \omega(I_0)\rangle = 0$ and $\mathrm{D}\omega(I_0)\,v = -\mu\,\omega(I_0)$.
- Then the existence theorem about (15.4.34) follows from the inversion theorem of multivariable calculus (see eg. HILDEBRANDT [Hil], Analysis 2, Chapter I.9). □

In BROER and HUITEMA [BH], it is shown that under the hypothesis (15.4.33), a statement analogous to Theorem 15.33 follows for the invariant tori in the energy surfaces $H_\varepsilon^{-1}(E)$ of the perturbed system.

15.38 Exercise (Nondegeneracy Conditions)
For the Hamiltonians $H_i : \mathbb{R}^2 \times \mathbb{T}^2 \to \mathbb{R}$,

$$H_1(I,\varphi) = I_1 + I_2 + I_1^2 \quad \text{and} \quad H_2(I,\varphi) = I_1 + I_2 + I_1^2 - I_2^2\,,$$

show: In the case of H_1, the invariant torus $\{0\} \times \mathbb{T}^2$ is (15.4.31)-degenerate, but is nondegenerate in the sense of condition (15.4.33). In the case of H_2, show the opposite.[15] ♢

15.39 Literature The book 'The KAM Story' by DUMAS [Dum] provides an overview over the history and ramifications of KAM theory. ♢

15.5 Diophantine Condition and Continued Fractions

We now consider the diophantine condition (15.3.8) for the simplest case of dimension $n = 2$. In this case, the set of independent frequency vectors is

$$\Omega_{\gamma,\tau} = \left\{ (\omega_1,\omega_2) \in \mathbb{R}^2 \;\middle|\; \forall (\ell_1,\ell_2) \in \mathbb{Z}^2 \backslash \{0\} : \; |\ell_1\omega_1 + \ell_2\omega_2| \ge \gamma\,(\ell_1^2 + \ell_2^2)^{-\tau/2} \right\}. \tag{15.5.1}$$

The defining condition is certainly violated when either $\omega_1 = 0$ or $\omega_2 = 0$. So we divide by ω_2 in (15.5.1) and let $\omega := \omega_1/\omega_2$. For $(\omega_1,\omega_2) \in \Omega_{\gamma,\tau}$, it is clear that ω is irrational and (after renaming $(q,p) := (\ell_1, -\ell_2)$ and redefining the constant γ) satisfies the inequalities

$$|q\omega - p| \ge \frac{\gamma}{(p^2+q^2)^{\tau/2}} \quad \text{resp.} \quad \left|\omega - \frac{p}{q}\right| \ge \frac{\gamma}{|q|(p^2+q^2)^{\tau/2}}\,. \tag{15.5.2}$$

These are most difficult to satisfy if we reduce p and q by their greatest common divisor. We also assume, with no loss of generality, that $q \in \mathbb{N}$. The diophantine condition for ω thus means poor approximability of ω by rational numbers p/q.

[15] The examples are taken from the article [Do] by R. DOUADY.

The two inequalities in (15.5.2) lead to different measures of approximability:

15.40 Definition
For $\omega \in \mathbb{R}$ we call the rational number p/q (written in lowest terms)

1. **a best approximant of the first kind for** ω *if*

$$\left|\omega - \frac{p}{q}\right| < \left|\omega - \frac{p'}{q'}\right|$$

for all $p'/q' \in \mathbb{Q}\backslash\{p/q\}$ with denominator $1 \le q' \le q$.

2. **a best approximant of the second kind for** ω *if*

$$|q\omega - p| < \left|q'\omega - p'\right|$$

for all $p'/q' \in \mathbb{Q}\backslash\{p/q\}$ with denominator $1 \le q' \le q$.

The second property implies the first, but not vice versa:

15.41 Example (Best Approximants)
For $\omega := 1/5$, the rational $p/q := 1/3$ is a best approximant of the first, but not of the second kind, because in both cases, only the fraction $p'/q' := 0/1$ needs to be considered, and it satisfies $\left|\omega - \frac{p}{q}\right| = 2/15 < 1/5 = \left|\omega - \frac{p'}{q'}\right|$, but $|q\omega - p| = |3/5 - 1| = 2/5 > 1/5 = \left|q'\omega - p'\right|$. ◊

It now transpires that best approximants of the second kind can be found by the continued fraction expansion of ω. To this end we assume $\omega > 0$ and consider in the quadrant $[0, \infty)^2 \subset \mathbb{R}^2$ the ray through the origin with slope ω.

This ray divides the quadrant, and also the set $\mathbb{N}_0^2 \subset [0, \infty)^2$ of lattice points, into

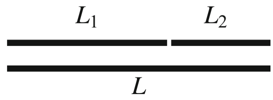
$$M_\pm := \left\{(k_1, k_2) \in \mathbb{N}_0^2 \mid \pm(k_2 - k_1\omega) \ge 0\right\}\backslash\{(0, 0)\}.$$

It is only in the case of rational slope $\omega \in \mathbb{Q}$ that these two sets fail to be disjoint. The convex hulls of $M_\pm$ have lattice points (q_n, p_n) as their extremal points, starting with $(q_0, p_0) := (1, \lfloor \omega \rfloor)$ for M_- and $(q_{-1}, p_{-1}) := (0, 1)$ for M_+.

These are best approximants of the second kind, because their vertical distance from the straight line is smaller than the distances of all lattice points (q', p') whose first coordinate q' is smaller.

15.42 Example (Golden Ratio g)
The following construction has been known since antiquity:

Divide a segment of length L into two parts of lengths $L_1 = g \cdot L$, $L_2 = L - L_1$ such that the thus obtained two ratios are equal: $\frac{L_1}{L} = \frac{L_2}{L_1}$.



One concludes that $g = (1-g)/g$, or $g^2 + g - 1 = 0$, hence[16] $g = \frac{\sqrt{5}-1}{2} \approx 0.618$. Here the extremal points are $(q_n, p_n) = (F_n, F_{n-1})$, see Figure 15.5.1, where (F_n) is the sequence of *Fibonacci numbers* defined by

$$F_{-1} := 1\,,\ F_0 := 0\,, \quad \text{and} \quad F_n := F_{n-2} + F_{n-1} \quad (n \in \mathbb{N})$$

(hence $F_1 = F_2 = 1$, $F_3 = 2$, $F_4 = 3$, $F_5 = 5$, $F_6 = 8$, $F_7 = 13$, $F_8 = 21 \ldots$; commonly one starts at F_1) .
In this case, the slopes $g_n := p_n/q_n$ of the rays through the origin and the extremal points converge towards the golden ratio g, because the iteration

$$g_{n+1} = \frac{F_n}{F_{n+1}} = \frac{F_n}{F_n + F_{n-1}} = \frac{1}{1+g_n}$$

has the stable fixed point $g = \frac{1}{1+g} = \frac{1}{1+\frac{1}{1+\frac{1}{1+\ldots}}}$:

$$g_1 = 0\,,\ g_2 = 1\,,\ g_3 = \tfrac{1}{2}\,,\ g_4 = \tfrac{2}{3} \approx 0.667\,,\ g_5 = \tfrac{3}{5} = 0.6\,,\ g_6 = \tfrac{5}{8} = 0.625\,, \ldots .$$

Generally, the lattice points (q_n, p_n) are calculated by a continued fraction expansion of $\omega \in \mathbb{R}^+$. ◊

15.43 Literature A good reference is KHINCHIN [Kh]. ◊

[16] Frequently, the reciprocal $1/g = g + 1$ is instead called the golden ratio.

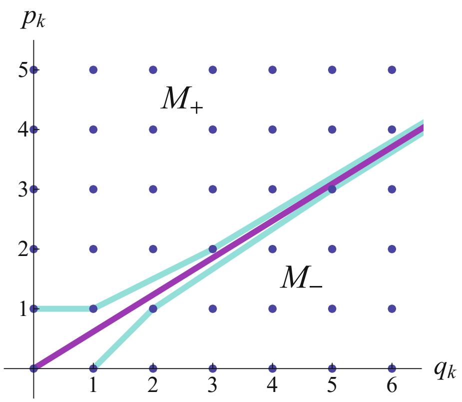
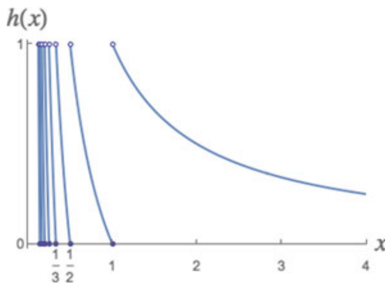


Figure 15.5.1 Ray with slope g and extremal points (q_n, p_n)

15.44 Definition

- *The* **Gauss map**[17] *is the piecewise continuous function (see figure)*

$$h : [0, \infty) \to [0, 1) \quad , \quad h(0) := 0,\ h(x) := 1/x - \lfloor 1/x \rfloor \text{ for } x > 0\,.$$



- *Assume (for simplicity) that ω is irrational. For the set*

$$\mathcal{F} := \{a : \mathbb{N}_0 \to \mathbb{Z} \mid \forall n \in \mathbb{N} : a_n \in \mathbb{N}\}$$

of integer sequences and $\Phi : [0, \infty) \to \mathcal{F}$,

$$\Phi(\omega)_0 := \lfloor \omega \rfloor$$
$$\Phi(\omega)_n := \lfloor 1/h^{(n-1)}(\{\omega\}) \rfloor \quad (n \in \mathbb{N})$$

(remember that $x = \lfloor x \rfloor + \{x\}$) we call $a := \Phi(\omega)$ the sequence of **partial denominators** *of ω.*

- *For the start values $\binom{q_{-2}}{p_{-2}} := \binom{1}{0}$, $\binom{q_{-1}}{p_{-1}} := \binom{0}{1}$ and the sequences $p, q : \mathbb{N}_0 \to \mathbb{N}_0$ defined by*

$$\binom{q_n}{p_n} := a_n \binom{q_{n-1}}{p_{n-1}} + \binom{q_{n-2}}{p_{n-2}} \tag{15.5.3}$$

(thus $\binom{q_0}{p_0} = \binom{1}{\lfloor \omega \rfloor}$), we call $\omega_n := \frac{p_n}{q_n}$ the nth **convergent** *of the continued fraction for ω.*

---

[17] In Exercise 9.6 we instead used the domain [0, 1).
Not to be confused with the mapping in differential geometry bearing the same name and defined on page 117.

15.45 Example (The Convergents for $\sqrt{5}$)
For $\omega := \sqrt{5} \approx 2.23607$, one has $\{\omega\} = \sqrt{5} - 2 = 1/(\sqrt{5}+2) \approx 0.23607$, thus $h^{(k)}(\{\omega\}) = \{\omega\} \quad (k \geq 1)$. Therefore $a_0 = 2$ and $a_\ell = 4$ ($\ell \geq 1$), hence $\omega_0 = 2$, $\omega_1 = 2 + \frac{1}{4} = \frac{9}{4} = 2.25$,

$$\omega_2 = 2 + \frac{1}{4+\frac{1}{4}} = \frac{38}{17} \approx 2.2353 \quad \text{and} \quad \omega_3 + \frac{1}{4+\frac{1}{4+\frac{1}{4}}} = \frac{161}{72} \approx 2.23611\,. \qquad \diamondsuit$$

15.46 Theorem (Continued Fractions)
The convergents $\omega_n = p_n/q_n$ of the irrational number ω, satisfy:

1. $\det\begin{pmatrix} q_{n-2} & q_{n-1} \\ p_{n-2} & p_{n-1}\end{pmatrix} = (-1)^n \quad (n \in \mathbb{N}_0)$, *hence* $\omega_{n-1} - \omega_n = \frac{(-1)^n}{q_{n-1}q_n}$.
2. *Numerator and denominator in the representation $\frac{p_n}{q_n}$ of ω_n are relatively prime.*
3. *The sequence $(\omega_{2m})_{m\in\mathbb{N}_0}$ is increasing, the sequence $(\omega_{2m+1})_{m\in\mathbb{N}_0}$ is decreasing.*
4. *The denominators of the convergents satisfy $q_n \geq 2^{(n-1)/2} \quad (n \geq 2)$.*

Proof:

1. Starting an induction with $\det\begin{pmatrix} q_{-2} & q_{-1} \\ p_{-2} & p_{-1}\end{pmatrix} = \det\begin{pmatrix} 1 & 0 \\ 0 & 1\end{pmatrix} = (-1)^0$ and (15.5.3), the claim follows, because the induction step is

$$\det\begin{pmatrix} q_{n-1} & q_n \\ p_{n-1} & p_n\end{pmatrix} = \det\begin{pmatrix} q_{n-1} & a_n q_{n-1}+q_{n-2} \\ p_{n-1} & a_n p_{n-1}+p_{n-2}\end{pmatrix} = \det\begin{pmatrix} q_{n-1} & q_{n-2} \\ p_{n-1} & p_{n-2}\end{pmatrix} = (-1)^{n+1}\,.$$

2. This follows from part 1 of the theorem, because a common divisor of p_n and q_n must also divide $(-1)^n$, as $p_{n-1}q_n - q_{n-1}p_n = (-1)^n$.
3. This follows from (15.5.3) and part 1 of the theorem, because

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{\det\begin{pmatrix} q_{n-2} & q_n \\ p_{n-2} & p_n\end{pmatrix}}{q_{n-2}\,q_n} = a_n \frac{\det\begin{pmatrix} q_{n-2} & q_{n-1} \\ p_{n-2} & p_{n-1}\end{pmatrix}}{q_{n-2}\,q_n} = (-1)^n \frac{a_n}{q_{n-2}\,q_n}\,.$$

4. This follows inductively, starting at $q_3 \geq q_2 = a_2 + 1 \geq 2$ and with induction step $q_n = a_n q_{n-1} + q_{n-2} \geq q_{n-1} + q_{n-2} \geq \frac{1}{2}\big(2^{n/2} + 2^{(n-1)/2}\big)$. □

15.47 Exercise (Continued Fraction Expansion) Show that for irrational ω, the continued fraction expansion converges, i.e., $\lim_{n\to\infty}\omega_n = \omega$. ◊

Together with Theorem 15.46, this implies an estimate for the speed of convergence:

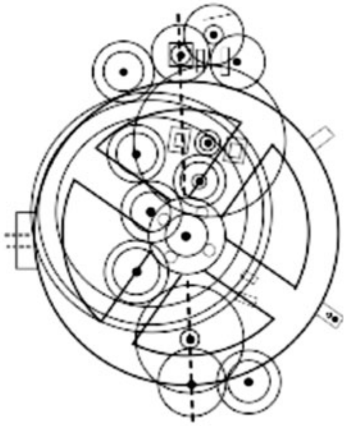
15.48 Corollary *All irrational numbers ω can be approximated by rational numbers in the following sense:*

$$\left|\omega - \frac{p_{n-1}}{q_{n-1}}\right| < \frac{1}{2q_{n-1}^2} \quad \textit{or} \quad \left|\omega - \frac{p_n}{q_n}\right| < \frac{1}{2q_n^2} \qquad (n \in \mathbb{N}).$$

Proof: By parts 3 and 1 of Theorem 15.46, and since $q_{n-1} < q_n$, one obtains

$$\left|\frac{p_{n-1}}{q_{n-1}} - \omega\right| + \left|\omega - \frac{p_n}{q_n}\right| = \left|\frac{p_{n-1}}{q_{n-1}} - \frac{p_n}{q_n}\right| = \frac{1}{q_{n-1}\,q_n} < \frac{1}{2}\left(\frac{1}{q_{n-1}^2} + \frac{1}{q_n^2}\right). \qquad \square$$

The Antikythera Mechanism
This calculational tool for astronomy was found in 1900 in a ship that had sunk in front of the coast of the Greek island Antikythera. It is from about the 2nd century BC and consists of more than 30 gears. It is the only known instrument of its kind from that era. (Left image: National Archaeological Museum, Athens (Greece) (NAM inv. No. X 15087); The rights of the depicted monuments belong to the Greek Ministry of Culture and Sports (Law 3028/2002). Right image: De Solla Price, courtesy of *Transactions of the American Philosophical Society*, Vol 64 No 7 (1974))
Tony Phillips speculates on the *Feature Column* webpage of the American Mathematical Society how the excellent mechanical approximation $254/19 = 13.368\,421...$ (the *metonic cycle*) to the astronomical ratio $13.368\,267...$ of year and sidereal month was found. Indeed, from the continued fraction expansion $13.368\,267.. = [13, 2, 1, 2, 1, 1, 17, ...]$, one obtains $[13, 2, 1, 2, 1, 1] = 254/19$, whereas the next better approximation $[13, 2, 1, 2, 1, 1, 17] = 4465/334 = 13.368\,263..$ could probably not have been realized mechanically.



15.49 Remark (Parameters in the Diophantine Condition)
We see from this that the exponent τ in the diophantine condition (15.5.1) must be larger or equal to 1, because otherwise, this condition could not be fulfilled for a single ω.

For $\tau = 1$, by Corollary 15.48, the constant γ must be $\leq \frac{1}{2}$ for the set $\Omega_{\gamma,\tau}$ to be non-empty. The example of the golden section $\omega = g$ shows that this estimate is realistic (see KHINCHIN [Kh], Chapter 7).

In a diophantine sense, g is the worst approximable number because the partial denominators of its continued fraction expansion are all 1, hence minimal. ◊

15.6 Cantori: In the Example of the Standard Map

The **standard map**, or **Chirikov-Taylor map** is a family of mappings $F_\varepsilon = (F_{\varepsilon,1}, F_{\varepsilon,2}) \in C^\infty(\mathbb{R}^2, \mathbb{R}^2)$ of the plane into itself, given by

$$F_\varepsilon(x, y) = \big(x + y + \varepsilon \sin(2\pi x)\,,\; y + \varepsilon \sin(2\pi x)\big)\,. \tag{15.6.1}$$

These maps satisfy the following elementary, but important, properties:

1. F_ε is *area preserving*, because $\mathrm{D}F_\varepsilon(x, y) = \begin{pmatrix} 1+2\pi\varepsilon\cos(2\pi x) & 1 \\ 2\pi\varepsilon\cos(2\pi x) & 1 \end{pmatrix}$.
2. F_ε is *doubly periodic*, i.e., $F_\varepsilon(x + \ell_x, y + \ell_y) = F_\varepsilon(x, y) + (\ell_x, \ell_y)$ for all $(x, y) \in \mathbb{R}^2$ and $(\ell_x, \ell_y) \in \mathbb{Z}^2$. Thus F_ε defines a family of area preserving mappings $f_\varepsilon : \mathbb{T}^2 \to \mathbb{T}^2$ of the 2-torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.
   We use $x, y \in [0, 1)$ as coordinates.
3. The mappings $f_\varepsilon = (f_{\varepsilon,1}, f_{\varepsilon,2})$ have the *twist property*

$$\frac{\partial f_{\varepsilon,1}}{\partial y}(x, y) > 0 \qquad \big((x, y) \in \mathbb{T}^2\big)\,.$$

The mapping f_0 is very simple, namely it is *integrable* in the following sense:

- The torus $\mathbb{T}^2$ is foliated into the f_0-invariant circles

$$S_r := \{(x, y) \in \mathbb{T}^2 \mid y = r\}\,,$$

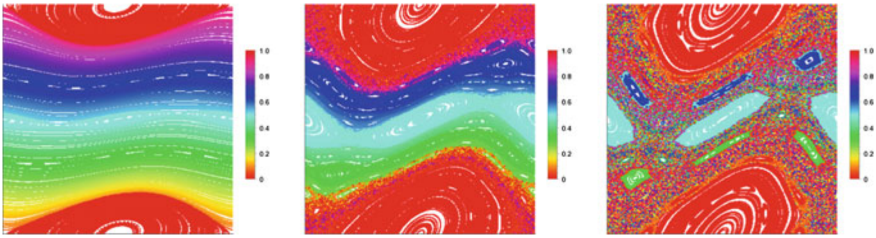
  which are parametrized by their rotation number $r \in [0, 1)$.
- On these 1-dimensional tori S_r, the mapping f_0 acts as a translation by r.

For $\varepsilon > 0$, the S_r are no longer f_ε-invariant. On the other hand, for f_0, the twist property is analogous to the condition (15.4.31) of independent variation of the frequencies. Indeed, there is an analog to the KAM theorem (Theorem 15.33), according to which for small ε, a large[18] subset M_ε of the phase space $\mathbb{T}^2$ is still foliated into f_ε-invariant circles.

As ε is increased, more and more of these f_ε-invariant tori (each identified by its rotation number) will be not just deformed, but destroyed, see Figure 15.6.1. For large values of ε, there are no more such tori, as is proved in [Mac] by MacKay and Percival.

The mappings f_ε have the *reflection symmetry* $f_\varepsilon \circ I = I \circ f_\varepsilon$ with

[18]The Haar measure of $\mathbb{T}^2 \backslash M_\varepsilon$ is of the order $\mathcal{O}(\sqrt{\varepsilon})$.



$\varepsilon = 0.05$ $\varepsilon = 0.155$ $\varepsilon = 0.19$

Figure 15.6.1 Numerical iterations of the standard map. Color coding according to the continued fraction $R_\varepsilon(z)$ of the initial value $z \in \mathbb{T}^2$

$$I : \mathbb{T}^2 \to \mathbb{T}^2 \quad , \quad (x, y) \mapsto (1 - x, 1 - y) \quad (\text{mod } 1).$$

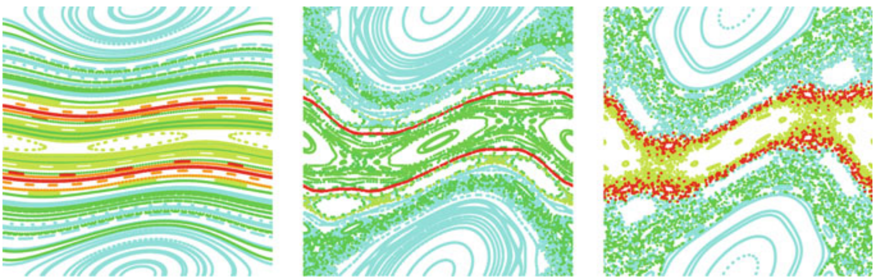
So I maps orbits into orbits and is therefore a symmetry of the phase space portraits (see Figure 15.6.1).

We can assign a rotation number to points $z := (x, y)$ in the phase space $\mathbb{T}^2$ even if $\varepsilon \neq 0$. This is done in analogy to (2.2.4) by

$$R_\varepsilon(z) := \lim_{n\to\infty} \frac{1}{n} F^{(n)}_{\varepsilon,1}(z) ,$$

where $F^{(n)}_\varepsilon$ is the nth iterate of F_ε. By Birkhoff's ergodic theorem (Theorem 9.32), this limit exists for almost all points in phase space, with respect to Haar measure. The coloring of Figure 15.6.1 at the point z indicates the value of $R_\varepsilon(z)$. Any two f_ε-invariant tori $S, S' \subset \mathbb{T}^2$ divide the phase space into the two connected components of $\mathbb{T}^2 \backslash (S \cup S')$. These are themselves invariant, and are annuli (that is, homeomorphic to the cartesian product of an open interval and S^1).

The last two surviving tori (up to approximately $\varepsilon = 0.155$) have the rotation numbers of the golden ratio, $g = \frac{1}{2}(\sqrt{5} - 1)$ and $1 - g$ respectively (Figure 15.6.2).



$\varepsilon = 0.05$ $\varepsilon = 0.155$ $\varepsilon = 0.19$

Figure 15.6.2 Numerical iterations of the standard map. Color coding according to similarity of the continued fraction of $R_\varepsilon(z)$ and the golden ratio

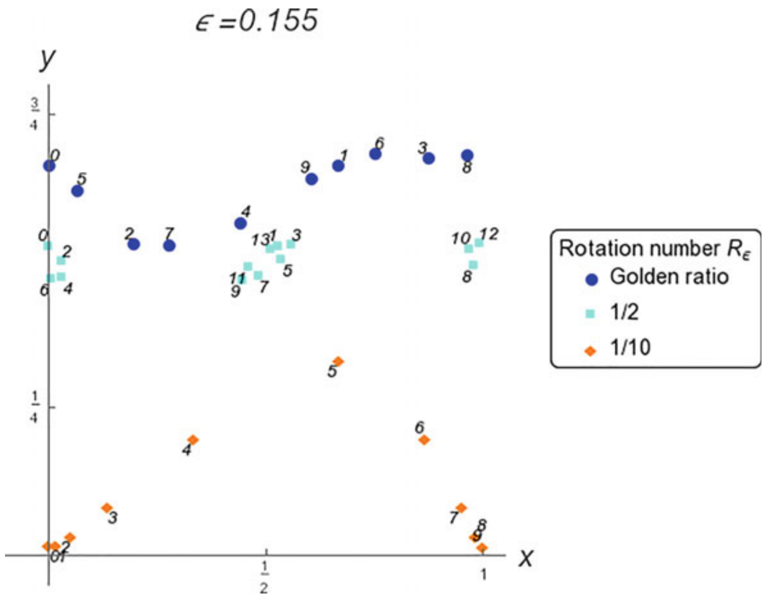


Figure 15.6.3 Iterations each under the standard map f_ε for $\varepsilon = 0.155$, for three rotation numbers

In Figure 15.6.3, one can lucidly discern the reason for the robustness of these two invariant tori. Namely, for the rotation number g, the iterates are distributed particularly evenly, and the orbit is therefore less influenced by the sine shaped perturbation in (15.6.1).

In the orbits shown in Figure 15.6.2, the color indicates the number of convergents in the continued fraction expansion of the (numerically determined) rotation number that coincide with those of the golden ratio g or $1 - g$. Red designates the orbits that match the golden ratio best. This visualizes the above statement about the last surviving invariant tori.

The maps f_ε have the obvious fixed points $(0, 0)$ and $(1/2, 0)$, whereas $(0, 1/2)$ and $(1/2, 1/2)$ form an orbit of period 2.

- As the linearization of f_ε at $(0, 0)$ has the form $\begin{pmatrix} 1+2\pi\varepsilon & 1 \\ 2\pi\varepsilon & 1 \end{pmatrix}$, this fixed point is hyperbolic for $\varepsilon > 0$.
- Conversely, the fixed point $(1/2, 0)$ is elliptic (for $\varepsilon > 0$ not too large), since its linearization equals $\begin{pmatrix} 1-2\pi\varepsilon & 1 \\ -2\pi\varepsilon & 1 \end{pmatrix}$.
- At the point $(0, 1/2)$ of the periodic orbit the linearization of $f_\varepsilon^{(2)}$ equals $\begin{pmatrix} 1+2\pi\varepsilon & 1 \\ 2\pi\varepsilon & 1 \end{pmatrix}$ $\begin{pmatrix} 1-2\pi\varepsilon & 1 \\ -2\pi\varepsilon & 1 \end{pmatrix}$. The trace equals $2 - (2\pi\varepsilon)^2$, so that this orbit is elliptic, too.

The rotation numbers near an elliptic periodic point are locally constant. This corresponds to 'islands' of uniform color in Figure 15.6.1 (right).[19]

[19] In Figure 15.6.1, for points near $(0, 1/2)$ the rotation number is equal to 0 (mod 1), whereas for points near $(1/2, 1/2)$ it equals $1/2$ (mod 1).

However, it is not obvious at all that for every given number r, there exists a f_ε-orbit with this rotation number r. This was shown by MATHER in [Mat], using the twist property and essentially relying on low dimensional properties, namely the order structure of $\mathbb{R}$. If r is rational, then there exists a periodic orbit with this rotation number.

For irrational r and large ε, the orbits with this rotation number form what is called a *cantorus*, rather than a KAM torus, namely an f_ε-invariant Cantor set.

15.50 Literature Surveys of this *Aubry-Mather theory* can be found in MOSER, [Mos5] and SIBURG, [Sib]. ◊

Chapter 16
Relativistic Mechanics

City of Tübingen, viewed at pedestrian speed (left) and at 4/5 of the speed of light (right) Picture: courtesy of Ute Kraus and Marc Borchers [KB], http://www.spacetimetravel.org.

The α principle of relativity states that in the laws of physics, only *relative* velocities occur, so that it is in particular meaningless to postulate a state of absolute rest.

The first *theory of relativity* (thus understood as a kinematic theory building on the principle of relativity) is due to Galileo Galilei. Similar to the special theory of relativity by Albert Einstein, it assumes the isotropy of the space $\mathbb{R}^3$ and homogeneity of the spacetime $\mathbb{R}^3 \times \mathbb{R}$ and states that laws of physics have the same form in all inertial systems (coordinate systems that are not accelerated). Galilei justifies his theory of relativity in his book *Dialogue Concerning the Two Chief World Systems* [1] [Gal1], which appeared in 1632. He argued by the impossibility to determine the speed of a ship by experiment from within a closed room on that ship.

If the relative velocity of one reference frame a with respect to another reference frame b is denoted as $v_{a,b} \in \mathbb{R}^3$, then according to Galilei, one obtains the rule of vector addition of the velocities,

[1] In which he promoted the 'Copernican World System' and which led to his persecution by the church.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_16

$$v_{1,3} = v_{1,2} + v_{2,3} \,. \tag{16.0.1}$$

The distinction to Einstein's special theory of relativity is Galilei's assumption that the speed of light is infinity.

It has however become common practice to refer to dynamical systems in which the occurring speeds are small compared to the speed of light as *nonrelativistic*, whereas the term *relativistic* is used for Einstein's theory of relativity.

16.1 The Speed of Light

"Therefore, special relativity differs from classical mechanics not by the postulate of relativity, but only by the postulate that the speed of light in vacuum is a constant." ALBERT EINSTEIN, in [Ei2]

The question whether light propagates with finite speed or instantaneously had been disputed since the days of the natural philosophers of Greek antiquity.

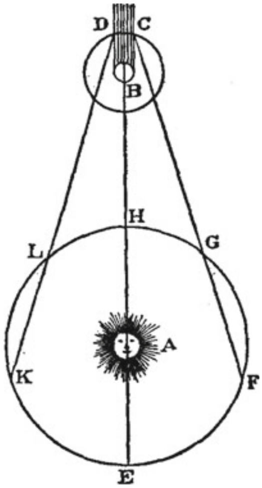
A Measurement of the Speed of Light
Salviati: "The experiment which I devised was as follows: Let each of two persons take a light contained in a lantern, or other receptacle, such that by the interposition of the hand, the one can shut off or admit the light to the vision of the other. Next let them stand opposite each other at a distance of a few cubits and practice until they acquire such skill in uncovering and occulting their lights that the instant one sees the light of his companion he will uncover his own. [...] Having acquired skill at this short distance let the two experimenters, equipped as before, take up positions separated by a distance of two or three miles and let them perform the same experiment at night, noting carefully whether the exposures and occultations occur in the same manner as at short distances; if they do, we may safely conclude that the propagation of light is instantaneous [...]
Sagredo: "This experiment strikes me as a clever and reliable invention. But tell us what you conclude from the results."
Salviati: "In fact I have tried the experiment only at a short distance, less than a mile, from which I have not been able to ascertain with certainty whether the appearance of the opposite light was instantaneous or not; but if not instantaneous it is extraordinarily rapid, I should call it momentary [...]"
GALILEO GALILEI:
Two New Sciences (1638), quoted after [Gal2], pages 363–364.

As illustrated in the boxed quotation, Galileo suggested to decide the question by measurement. Galileo's method only provided a lower bound for the speed of light, but a different suggestion was to lead to success shortly thereafter.

A key problem for nautical navigation at that time was to determine the geographical longitude. It is easy to determine the latitude, say, by measuring the altitude of the polar star above the horizon. In contrast, in order to determine the longitude by measuring the altitude of a star, or the sun, above the western or eastern horizon, knowledge of the exact time is required.

However, the clocks available on ships at that time were not of sufficient precision, in particular over long journeys. Galileo suggested to use the pre-calculated eclipses of Jupiter's moons as a universal clock. As a matter of fact, the method was not practical to use on the open sea (see the fascinating book *Longitude* by DAVA SOBEL [Sob1]), but it was used on land.

In the year 1671, the Danish astronomer Ole Rømer visited the island Hven in order to determine the latitude of Tycho Brahe's old observatory and thus to connect his data with the data of the observatory in Paris.



Rømer's sketch of Earth, Sun, Jupiter, and Io.

During a series of observations over about 8 months, including 140 eclipses of Jupiter's moon Io, he observed a deviation from the calculations that varied with the distance between Jupiter and earth.

In 1676, he published his theory that these deviations are due to the finiteness of the speed of light (see Ole Rømer's sketch, from [Roe]), and he estimated the time that light takes to travel from the sun to the earth to be about 11 min. (For the history of measuring the speed of light, see also MACKAY and OLDFORD in [MKO]).

This was a refutation of Galilei's *theory* of relativity, but not of his *principle* of relativity.

Now Huygens and others argued that no material particles could be that fast. Instead, one should imagine the propagation of light like a density variation in some substance filling the space, similar to the propagation of sound in air.

This *ether theory* was now in contradiction to the invariance principle by Galilei, and later, Newton, namely the principle that only relative velocities, but not absolute velocities, can be defined.

Actually, it transpired in 1887 during a more precise measurement of the speed of light by Michelson and Morley, that the speed of light was not affected by the motion of the earth around the sun. Thus a new theory of relativity was needed, since both Galilei's theory of relativity and ether theory had been refuted. In 1905, EINSTEIN succeeded to do this in his article *Zur Elektrodynamik bewegter Körper* (On the electrodynamics of moving bodies) [Ei1].

Since 1983, one meter has been defined as that distance that light traverses in vacuum during the time span of one 299 792 458th of a second. So the speed of light in units of meter per second can no longer be measured; instead one would be measuring the length of the 'international prototype metre' stored in Paris, in terms of the newly defined meter.

We will proceed to use units in which the speed of light is 1. In order to interpret formulas in a different system of units, all occuring velocities (in that system of units) have to be divided by the speed of light c.

16.2 The Lorentz- and Poincaré Groups

"Es ist eine Mannigfaltigkeit und in derselben eine Transformationsgruppe gegeben; man soll die der Mannigfaltigkeit angehörigen Gebilde hinsichtlich solcher Eigenschaften untersuchen, die durch die Transformationen der Gruppe nicht geändert werden." FELIX KLEIN (1872), in [Kl] [2]

From a mathematical point of view, Einstein's special theory of relativity consists basically of the theory of the *Poincaré group* and a subgroup, the *Lorentz group*.

We begin by studying the Lorentz group (as a special case of groups called indefinite-orthogonal), and then explore its relevance in physics. This follows the approach in the *Erlangen Program* by Felix Klein, according to which geometry is analyzed by investigating its transformation group.

The geometry itself is defined by a bilinear form on spacetime:

16.1 Definition

- *For* $m, n \in \mathbb{N}_0$ *and* $k := m + n$*, define a symmetric bilinear form with* **signature** (m, n) *by*

$$\langle \cdot, \cdot \rangle_{m,n} : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R} \quad , \quad \langle v, w \rangle_{m,n} = \sum_{\ell=1}^{m} v_\ell w_\ell - \sum_{\ell=m+1}^{k} v_\ell w_\ell \, . \qquad (16.2.1)$$

- *The* **indefinite orthogonal group** *is the transformation group of this bilinear form:*

$$\mathrm{O}(m, n) := \left\{ A \in \mathrm{GL}(k, \mathbb{R}) \mid \forall v, w \in \mathbb{R}^k : \langle Av, Aw \rangle_{m,n} = \langle v, w \rangle_{m,n} \right\} .$$

- $\left(\mathbb{R}^4, \langle \cdot, \cdot \rangle_{3,1}\right)$ *is called* **Minkowski space**. $\mathrm{L} := \mathrm{O}(3, 1)$ *is also known as the* **Lorentz group**, *its elements are the* **Lorentz transformations**.

[2] Translation [M. W. Haskell, 1892]: "Given a manifoldness and a group of transformations of the same; to investigate the configurations belonging to the manifoldness with regard to such properties as are not altered by the transformations of the group."

- *The* **Poincaré group** *is the semidirect product* $P := \mathbb{R}^4 \rtimes \mathrm{L}$ *with the operation*

$$(a, A) \circ (b, B) := (a + Ab, AB) .$$

16.2 Remarks

1. By Sylvester's Law of Inertia (Theorem 6.13.1), every nondegenerate symmetric bilinear form on a k-dimensional real vector space has, with respect to an appropriate basis, the form (16.2.1) (and the numbers m and n do not depend on the choice of basis).
2. Due to $\langle \cdot, \cdot \rangle_{n,m} = -\langle M \cdot, M \cdot \rangle_{m,n}$, with the permutation matrix $M := \begin{pmatrix} 0 & 1\!\!1_m \\ 1\!\!1_n & 0 \end{pmatrix} \in \mathrm{Mat}(k, \mathbb{R})$, we have an isomorphism between $\mathrm{O}(n, m)$ and $\mathrm{O}(m, n)$.
3. $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{k,0}$ is the Euclidean scalar product, and $\mathrm{O}(k, 0) = \mathrm{O}(k)$ is the orthogonal group.
4. In terms of the diagonal matrix $I := \begin{pmatrix} 1\!\!1_m & 0 \\ 0 & -1\!\!1_n \end{pmatrix} = 1\!\!1_m \oplus (-1\!\!1_n) \in \mathrm{Mat}(k, \mathbb{R})$, one has $\langle \cdot, \cdot \rangle_{m,n} = \langle \cdot, I \cdot \rangle$. Consequently, a matrix $A \in \mathrm{Mat}(k, \mathbb{R})$ is in $\mathrm{O}(m, n)$ if and only if $A^\top I A = I$.
5. It is customary to number the components of a vector v in Minkowski space as $\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} \in \mathbb{R}^4$. Then v_4 is called its time component and $\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \in \mathbb{R}^3$ its space component. But one also encounters the notation $\begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix} \in \mathbb{R}^4$ with the bilinear form $\langle v, w \rangle_{1,3} = v_0 w_0 - v_1 w_1 - v_2 w_2 - v_3 w_3$, in which the 0th component is interpreted as time.
   In the physics literature, one writes briefly $v_\alpha w^\alpha$ instead of $\langle v, w \rangle_{3,1}$, with Einstein's summation convention being used. Latin (rather than Greek) indices of summation will then refer to only the spatial components, as for instance in

$$v_a w^a = v_1 w_1 + v_2 w_2 + v_3 w_3 . \qquad \diamondsuit$$

16.3 Theorem (Lorentz Group)

- *The indefinite orthogonal groups* $\mathrm{O}(m, n)$ *are Lie groups of dimension* $\frac{1}{2}k(k-1)$, *with* $k = m + n$. *In particular,* $\dim(\mathrm{L}) = 6$.
- *The polar decomposition of a matrix* $A \in \mathrm{O}(m, n)$ *is of the form* $A = OP$ *with*

$$O = \begin{pmatrix} O_m & 0 \\ 0 & O_n \end{pmatrix} \in \mathrm{O}(m) \times \mathrm{O}(n) \subset \mathrm{O}(m, n) ,$$

  and $P = \exp\left(\begin{pmatrix} 0 & \theta \\ \theta^\top & 0 \end{pmatrix}\right)$ *for* $\theta \in \mathrm{Mat}(n \times m, \mathbb{R})$.
- *If* $0 < m < k$, *then* $\mathrm{O}(m, n)$ *is not compact and has four connected components. The connected component containing the identity has a polar decomposition with* $O_m \in \mathrm{SO}(m)$ *and* $O_n \in \mathrm{SO}(n)$. *Two of the connected components have determinant* 1, *the other two have determinant* -1.

Proof:
• $\mathrm{O}(m,n)$ is a group, and it is the pre-image of the regular value $I = \begin{pmatrix} \mathbb{1}_m & 0 \\ 0 & -\mathbb{1}_n \end{pmatrix}$ of the mapping

$$\mathrm{GL}(k,\mathbb{R}) \longrightarrow \mathrm{Sym}(k,\mathbb{R}) \quad , \quad A \longmapsto A^\top I A$$

(compare with the special case of the orthogonal group $\mathrm{O}(k)$ in Example E.19.2). Thus $\mathrm{O}(m,n)$ is a Lie group, and

$$\dim\bigl(\mathrm{O}(m,n)\bigr) = \dim\bigl(\mathrm{GL}(k,\mathbb{R})\bigr) - \dim\bigl(\mathrm{Sym}(k,\mathbb{R})\bigr) = \binom{k}{2}.$$

• Note first that, along with A, its transpose $A^\top = IA^{-1}I$ is also in $\mathrm{O}(m,n)$, because $A^{-1} \in \mathrm{O}(m,n)$, and the involution $B \mapsto IBI$ maps $\mathrm{O}(m,n)$ to itself. Consequently, along with A, the positive matrix $A^\top A$ is also in $\mathrm{O}(m,n)$, and the same applies to its root

$$P = \bigl(A^\top A\bigr)^{1/2} = \exp(\tilde{\theta}) \quad \text{with} \quad \tilde{\theta} := \tfrac{1}{2}\log\bigl(A^\top A\bigr) \ \in \ \mathrm{Sym}(k,\mathbb{R}) \cap \mathfrak{o}(m,n)$$

in the polar decomposition $A = OP$. As the elements a of the Lie algebra $\mathfrak{o}(m,n) \subset \mathrm{Mat}(k,\mathbb{R})$ of $\mathrm{O}(m,n)$ are characterized by the relation $a^\top I + Ia = 0$, $\tilde{\theta}$ is of the form $\begin{pmatrix} 0 & \theta \\ \theta^\top & 0 \end{pmatrix}$ as stated.
• The number of connected components, and their nature, follows from the polar decomposition and the fact (proved in E.19.2) that the orthogonal group $\mathrm{O}(\ell)$ has the rotation group $\mathrm{SO}(\ell)$ and its composition with reflections of $\mathbb{R}^\ell$ in hyperplanes as its two connected components.

The values of $\det(A)$ are obtained from the polar decomposition $A = O\exp\bigl(\tilde{\theta}\bigr)$ by means of Theorem 4.12:

$$\det(A) = \det(O)\exp\bigl(\mathrm{tr}(\tilde{\theta})\bigr) = \det(O) = \det(O_m)\,\det(O_n)\,. \qquad \square$$

16.4 Remarks (Subgroups of the Lorentz Group)

1. The *indefinite special orthogonal group* $\mathrm{SO}(m,n)$ consists of the elements of $\mathrm{O}(m,n)$ with determinant $+1$, and thus, provided $m,n > 0$, it has exactly two connected components.
2. In the case of the Lorentz group $\mathrm{L} = \mathrm{O}(3,1)$, the linear mappings defined by $I = \mathrm{diag}(1,1,1,-1)$ and $-I = \mathrm{diag}(-1,-1,-1,1)$ are called *time reversal* and *space reflection* respectively, and the connected component of the identity element is called the *proper orthochronous* or *restricted Lorentz group*

$$\mathrm{L}_+^\uparrow := \mathrm{SO}^+(3,1)\,.$$

   Together with the component $\mathrm{L}_-^\uparrow := -I\mathrm{L}_+^\uparrow$, it forms the *orthochronous Lorentz group* $\mathrm{L}^\uparrow := \mathrm{L}_+^\uparrow \dot\cup \mathrm{L}_-^\uparrow$.

The other connected components are $L_+^\downarrow := -L_+^\uparrow$ and $L_-^\downarrow := I L_+^\uparrow$. Thus the lower index denotes the sign of the determinant.

3. For all $0 \le m' \le m$ and $0 \le n' \le n$, $O(m,n)$ contains subgroups that are isomorphic to $O(m',n')$, namely for example, to $A' \in O(m',n')$, one can assign $A := \mathbb{1}_{m-m'} \oplus A' \oplus \mathbb{1}_{n-n'} \in O(m,n)$.
   Hence the group $O(m) \times O(n)$ is also a subgroup of $O(m,n)$. In the case of the restricted Lorentz group, $SO^+(3,1) \cap \big(O(3) \times O(1)\big)$ consists of the rotations of space and is therefore a subgroup isomorphic to $SO(3)$. This symmetry corresponds to the property called *isotropy* of spacetime.
4. In contrast, the positive matrices in the polar decomposition (called the *special Lorentz transformations* or *Lorentz boosts*) do not from a subgroup of the Lorentz group, because this subset is not closed under multiplication (see (16.2.6)). ♢

16.5 Example (Two Dimensional Spacetime)
According to item 3 in the preceding remarks, one can view $O(1,1)$ as a subgroup of the Lorentz group $O(3,1)$. Within it, the connected component of 1 is[3]

$$SO^+(1,1) := \left\{ \exp\begin{pmatrix} 0 & \theta \\ \theta & 0 \end{pmatrix} = \begin{pmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{pmatrix} \;\middle|\; \theta \in \mathbb{R} \right\}, \tag{16.2.2}$$

so it is a subgroup of the restricted Lorentz group $SO^+(3,1)$.

The parameter θ is called *rapidity*. In comparison to the parametrization

$$\begin{pmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{pmatrix} = \begin{pmatrix} cc\frac{1}{\sqrt{1-v^2}} & \frac{v}{\sqrt{1-v^2}} \\ \frac{v}{\sqrt{1-v^2}} & \frac{1}{\sqrt{1-v^2}} \end{pmatrix} \tag{16.2.3}$$

in terms of the velocity $v = \tanh(\theta) \in (-1,1)$, rapidity has the advantage of additivity, because

$$\begin{pmatrix} \cosh(\theta_1) & \sinh(\theta_1) \\ \sinh(\theta_1) & \cosh(\theta_1) \end{pmatrix} \begin{pmatrix} \cosh(\theta_2) & \sinh(\theta_2) \\ \sinh(\theta_2) & \cosh(\theta_2) \end{pmatrix} = \begin{pmatrix} \cosh(\theta_1+\theta_2) & \sinh(\theta_1+\theta_2) \\ \sinh(\theta_1+\theta_2) & \cosh(\theta_1+\theta_2) \end{pmatrix}.$$

In contrast, the relativistic sum of the (parallel) velocities $v_i := \tanh(\theta_i)$ equals

$$v := \tanh\big(\operatorname{artanh}(v_1) + \operatorname{artanh}(v_2)\big) = \frac{v_1 + v_2}{1 + v_1 v_2} \in (-1,1)\,. \tag{16.2.4}$$

Thus, by adding these velocities relativistically, the speed of light will not be surpassed. Combining parallel velocities in this way is a commutative and associative operation, with the relativistic sum of three velocities being

$$\frac{v_1 + v_2 + v_3 + v_1 v_2 v_3}{1 + v_1 v_2 + v_1 v_3 + v_2 v_3}\,.$$

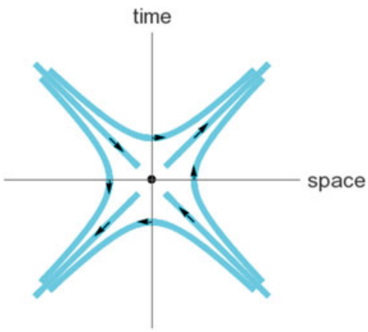
---

[3] Using the matrix exponential from (4.1.5).

For nonrelativistic velocities $|v_i| \ll 1$, one has in good approximation the same sum as for Galilei (see (16.0.1))

$$v \approx v_1 + v_2 \,.$$

The orbits of $SO^+(1,1)$ in $\mathbb{R}^2$ are (apart from the origin) hyperbolas and the half diagonals, see the figure on the right. ♢



We now transfer the parametrization (16.2.3) of the group $SO^+(1,1)$, which is isomorphic to $(\mathbb{R},+)$, to the six dimensional Lie group $SO^+(3,1)$.

16.6 Exercises (Lorentz Boosts)

1. Show that the Lorentz boosts, other than the identity $L(0) := 1\!\!1_4$, in the restricted Lorentz group $SO^+(3,1)$ are the following symmetric matrices:

$$L(v) := \gamma(v) \begin{pmatrix} (1\!\!1_3 - P_v)/\gamma(v) \; + \; P_v & v \\ v^\top & 1 \end{pmatrix} \qquad \big(v \in \mathbb{R}^3,\; 0 < \|v\| < 1\big), \tag{16.2.5}$$

with $\gamma(v) := (1 - \|v\|^2)^{-1/2}$ and the orthonormal projection $P_v = \frac{v \otimes v^\top}{\|v\|^2}$ on $\mathrm{span}(v) \subset \mathbb{R}^3$. (16.2.5) is the analog of (16.2.3).

2. Conclude for the polar decomposition $A = \tilde{O} P$ of a matrix $A = \left(\begin{smallmatrix} a & b \\ c^\top & d \end{smallmatrix}\right) \in SO^+(3,1)$ (with $a \in \mathrm{Mat}(3,\mathbb{R})$, $b, c \in \mathbb{R}^3$ and $d \in \mathbb{R}$) that the positive matrix is of the form $P = L(c/d)$, and therefore the orthogonal matrix is of the form $\tilde{O} = O \oplus 1 = A L(-c/d)$. Conclude further that $d = \sqrt{1 + \|b\|^2}$ and $b = Oc$. Thus the polar decomposition of A can be read off easily from the components of A.

3. Show, by using (16.2.5) and part 2, that the 'relativistic sum' of the velocities $v_1, v_2 \in \mathbb{R}^3$ with $\|v_i\| < 1$ is given by the formula

$$L(v_1)L(v_2) = \tilde{D}\, L(u) \quad \text{with} \quad u = \frac{v_1 + v_2 + \frac{v_1 \times (v_1 \times v_2)}{1 + \sqrt{1 - \|v_1\|^2}}}{1 + \langle v_1, v_2 \rangle} \tag{16.2.6}$$

and $\tilde{D} := D \oplus 1,\; D \in SO(3)$. This composition of velocities is neither commutative nor associative when $v_1,\; v_2$ are linearly independent. But check that at least the *norm* of u is a symmetric expression in the velocities v_1 and v_2:

$$\|u\|^2 = \frac{\|v_1 + v_2\|^2 - \|v_1 \times v_2\|^2}{(1 + \langle v_1, v_2 \rangle)^2} = 1 - \frac{\big(1 - \|v_1\|^2\big)\big(1 - \|v_2\|^2\big)}{(1 + \langle v_1, v_2 \rangle)^2} < 1 \,.$$

4. We write the composition (16.2.6) of Lorentz boosts in the form

$$L(v_1)L(v_2) = \tilde{D}_{12}\, L(u_{12}),\ \text{and analogously}\ L(v_2)L(v_1) = \tilde{D}_{21}\, L(u_{21})\,. \tag{16.2.7}$$

Conclude

- from the transformation law for Lorentz boosts (16.2.5) for velocities $v \in \mathbb{R}^3$ with $\|v\| < 1$ under space rotations $D \in \mathrm{SO}(3)$, namely

$$\tilde{D}L(v)\tilde{D}^\top = L(Dv) \qquad \text{with } \tilde{D} := D \oplus 1\,, \tag{16.2.8}$$

- from the invariance of the vector $(v_1 \times v_2) \oplus 0 \in \mathbb{R}^4$ under $L(v_1)$ and $L(v_2)$,
- and from the relation $L(v_2)L(v_1) = L(u_{12})\, \tilde{D}_{12}^\top$, which is the transpose of (16.2.7),

that $D_{12} = D_{21}^\top \in \mathrm{SO}(3)$ is a rotation matrix effecting a rotation about the axis spanned by $v_1 \times v_2$ and transforming u_{12} into u_{21}.

This rotation matrix is also called *Thomas matrix*.[4]

In UNGAR [Un], one can find a discussion of the Thomas rotation. ♢

16.3 Geometry of Minkowski Space

"Time flies like an arrow; fruit flies like a banana." attributed to Groucho Marx

We now know enough about the transformation group of Minkowski space to study its geometry in more detail.

16.7 Definition

1. *A vector u in Minkowski space $\left(\mathbb{R}^4, \langle\cdot,\cdot\rangle_{3,1}\right)$ is called* **timelike**, **lightlike**[5] *or* **spacelike**, *depending on whether $\langle u, u\rangle_{3,1} < 0$, $= 0$, or > 0 respectively.*
2. *The set $\mathcal{C} \subset \mathbb{R}^4$ of lightlike vectors is called the* **light cone**, *with the* **forward** *and* **backward light cone** *$\mathcal{C}^\pm := \{u \in \mathcal{C} \mid \pm u_4 > 0\}$ respectively.*

16.8 Exercise (Minkowski Product) Show that a pair v, w of timelike vectors always has a nonzero Minkowski product, $\langle v, w\rangle_{3,1} \neq 0$, but that the same is not necessarily true for pairs of spacelike vectors. ♢

16.9 Remarks (Speed of Light)

1. To a vector $u = (u_1, u_2, u_3, u_4)^\top \in \mathbb{R}^4$ with nonvanishing time component $u_4 \neq 0$, we attribute a velocity vector

---

[4] Named after the British mathematician and physicist *Llewellyn Hilleth Thomas* (1903–1992), who in 1926 predicted the relativistic precession of electrons in an atom that was then named after him.

[5] The assignment of the origin of $\mathbb{R}^4$ is not uniform over the literature.

$$V(u) := \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \Big/ u_4 \;\in\; \mathbb{R}^3$$

as in the nonrelativistic theory. The lightlike u have speed $\|V(u)\| = 1$, namely the speed of light. For timelike u, one has $\|V(u)\| < 1$.

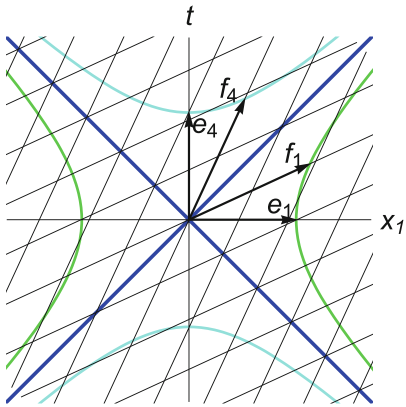
2. Lorentz transformations do not change the character of a vector in the sense of Definition 16.7.1, and they map the light cone $\mathcal{C}$ onto itself. The orthochronous Lorentz transformations (defined in Remark 16.4.2) map $\mathcal{C}^+$ and $\mathcal{C}^-$ each onto itself. ◊

Every basis of Minkowski space $\mathbb{R}^4$ defines a coordinate system for spacetime. But not every basis is appropriate for the purposes of physics. In any case, the canonical basis e_1, e_2, e_3, e_4 is a reasonable choice, because

$$\langle e_i, e_k\rangle_{3,1} = \delta_{i,k} s_k \quad \text{with} \quad s_1 = s_2 = s_3 = 1 \text{ and } s_4 = -1\,.$$

So the scales in space and time directions are normed.

The same applies for a basis f_1, f_2, f_3, f_4 if and only if there exists a Lorentz transformation M such that $f_k = M e_k$.
By Remark 16.9.2, f_1, f_2, f_3 will then be spacelike and f_4 timelike, see the figure on the right. It shows the hyperboloids consisting of those points u for which $\langle u, u\rangle_{3,1} = \pm 1$, as well as a coordinate net determined by the f-basis.



Conversely, if we only prescribe a velocity vector $v \in \mathbb{R}^3$, $\|v\| < 1$, this describes a Lorentz boost $L(v)$. It does not yet determine a unique Lorentz transformation M (and thus a coordinate system), but at least it determines *two equivalence relations* on spacetime:
For all inertial systems that move with the given velocity v with respect to the inertial system determined by the canonical basis e_1, e_2, e_3, e_4, these equivalence relations describe when two events take place at the *same location* or at the *same time* respectively.

Namely, if we denote, for a given timelike vector $f \in V^{-1}(v)$, the Lorentz-orthogonal subspace

$$f^\perp := \{e \in \mathbb{R}^4 \mid \langle e, f\rangle_{3,1} = 0\}\,,$$

then this space, as well as $\mathrm{span}(f)$, depend only on v.

- $f^\perp$ is 3-dimensional, and a minor extension of the statement from Exercise 16.8 shows that all vectors from $f^\perp \setminus \{0\}$ are spacelike. For an observer with the given velocity v, all events in an affine subspace $f^\perp + a$ (thus shifted by $a \in \mathbb{R}^4$) occur at the same time.

- In any affine subspace $\operatorname{span}(f) + a$, i.e., shifted by $a \in \mathbb{R}^4$, the events on this one-dimensional space take place at the same location.

16.10 Remark (Constant Speed of Light and the Lorentz Group)
Remark 16.9.2 justifies to some extent that the Lorentz transformations are the appropriate transformations of spacetime, because they leave the speed of light unchanged. While*dilatations*

$$\mathbb{R}^4 \longrightarrow \mathbb{R}^4 \quad , \quad x \longmapsto \lambda x \qquad \big(\lambda \in (0, \infty)\big) \tag{16.3.1}$$

also map the light cone to itself, they change the Minkowski bilinear form $\langle \cdot, \cdot \rangle_{3,1}$ by a factor λ^2.

For three space dimensions (but not for just one!), the extension of the orthochronous (see Remark 16.4) Poincaré group by the dilations (16.3.1) is the largest automorphism group of spacetime that preserves causality, see ZEEMAN [Zee]. It is in this sense that the quote by Einstein on page 442 can be understood. ◊

Thus the significance of the Minkowski bilinear form goes beyond defining the light cone (as well as future and past). We define the so-called *Minkowski norm*[6]

$$D(u) := \sqrt{|\,\langle u, u \rangle_{3,1}\,|} \qquad \big(u \in \mathbb{R}^4\big). \tag{16.3.2}$$

For a vector $u \in \mathbb{R}^3 \subset \mathbb{R}^4$ (hence with $u_4 = 0$), this quantity $D(u)$ does coincide with the Euclidean length. Likewise, for $u \in \mathbb{R}^1 \subset \mathbb{R}^4$ (hence $u = (0, 0, 0, u_4)^\top \in \mathbb{R}^4$), this same $D(u)$ measures the distance in time. So the Minkowski norm can be used to measure distances in space and in time.

As usual, the length of $u \in \mathbb{R}^3 \subset \mathbb{R}^4$ does not change under rotations ($D(Ou) = D(u)$ for $O \in SO(3)$). But the key point is that these quantities do not change under Lorentz boosts either, i.e., $D\big(L(v)u\big) = D(u)$.

In contrast, the norm $\sqrt{u_1^2 + u_2^2 + u_3^2}$ of the spatial component of u does change under Lorentz boosts:

16.11 Example (Lorentz Contraction)
As can be seen from the form (16.2.5) of a Lorentz boost with velocity $v \in \mathbb{R}^3$, $0 < \|v\| < 1$, one finds that lengths in the direction of the motion get shortened by a factor

$$\left\| \big((\mathbb{1}_3 - P_v) + \gamma(v) P_v\big)\; P_v \right\| = \gamma(v) = (1 - \|v\|^2)^{-1/2} < 1$$

where P_v denotes the (Minkowski orthogonal) projection onto the direction of the velocity. This phenomenon is called *Lorentz contraction*.

Orthogonal to the motion however, the same lengths are measured in both reference frames, because $\big((\mathbb{1}_3 - P_v) + \gamma(v) P_v\big)\; (\mathbb{1}_3 - P_v) = \mathbb{1}_3 - P_v$.

---

[6]Which is *not* a norm on $\mathbb{R}^4$!

The cause of this Lorentz contraction is that the notion called ‘simultaneous’ depends on the frame of reference.

For instance, if we want to determine the length of a rod passing by from left to right at the space origin of our reference frame, we measure the locations of both ends of the rod simultaneously in our reference frame, and take the difference.

In the reference frame of the rod, however, the measurement of the left end of the rod takes place at a later time than the measurement of the right end (which passes the origin first, in either reference frame), and this gives rise to an apparent shortening of its length (see Figure 16.3.1). ♢

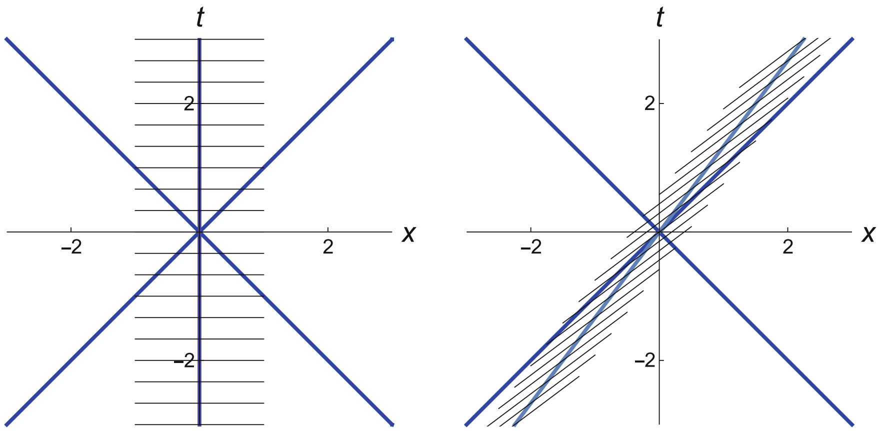


Figure 16.3.1 Lorentz contraction of a rod of length 2. Left: velocity 0; right: velocity $0.75\,c$

In a similar manner, a Lorentz boost also leads to a time contraction.

We call a curve $c : I \to \mathbb{R}^4$ in Minkowski space a *worldline*, and *timelike*[7] if it is regular and with timelike tangent vectors $c'(t)$ $(t \in I)$.

The *proper time* of a world line c is the time measured in the system that is parametrized by c.

16.12 Theorem

The proper time that elapses along a timelike world line $c \in C^1\big([t_0, t_1], \mathbb{R}^4\big)$ is

$$\tau(c) := \int_{t_0}^{t_1} D\big(c'(s)\big)\, \mathrm{d}s\,,$$

(with the Minkowski norm D from (16.3.2)). This proper time does not depend on the parametrization of the world line.

[7]Frequently, the requirement of being timelike is assumed as part of the definition of a worldline.

Proof:
The function $\varphi : [t_0, t_1] \to \mathbb{R}$, $\varphi(t) := \int_{t_0}^{t} D\big(c'(s)\big)\,\mathrm{d}s$ is continuously differentiable with derivative $\varphi'(t) = D\big(c'(t)\big) > 0$. Hence φ is invertible on the interval $[0, \tau] := \varphi\big([t_0, t_1]\big)$. The reparametrized worldline $\tilde{c} := c \circ \varphi^{-1} \in C^1\big([0, \tau], \mathbb{R}^4\big)$ is not only timelike, but also satisfies $D\big(\tilde{c}'\big) = 1$. Therefore the proper time is $\tau = \tau(\tilde{c}) = \tau(c)$. □

If the worldline c is parametrized by the time parameter t of an inertial system, i.e.,

$$c(t) = \begin{pmatrix} \hat{c}(t) \\ t \end{pmatrix} \quad \text{with} \quad \hat{c}(t) = \hat{c}(t_0) + \int_{t_0}^{t} v(s) \quad \mathrm{d}s$$

and with a velocity $v(s)$ at the time s of the inertial system, it has therefore the proper time

$$\tau(c) = \int_{t_0}^{t_1} \sqrt{1 - \|v(s)\|^2}\,\mathrm{d}s\,. \tag{16.3.3}$$

16.13 Example (The Hafele-Keating Experiment)
In an experiment by HAFELE and KEATING [HK] in 1971, four caesium atomic clocks were taken along flights around the earth, both in westbound and in eastbound direction. After the westbound flight, the clocks were an average of 273 ± 7 ns fast, compared to atomic clocks in Washington; after the eastbound flight, they were an average of 59 ± 10 ns slow compared to the same reference. The prediction of the theory of relativity derived from the flight data was that the clocks should differ by +275 ±21 and −40 ±23 ns respectively. This prediction is an additive combination of effects from special and general theory of relativity.

The experiment was repeated with some modifications in 2005, and the measurement agreed with the prediction with a relative precision of about 2 %. As for the effect from general relativity alone, it has in the meanwhile been confirmed to even a precision of 10^{-8} (see MÜLLER, PETERS and CHU [MPC]), but this was done with single atoms and a flight altitude of 0.1 mm.

Let us calculate the special relativity effect in an idealized variant of the experiment, in which the airplane flies along the equator with a constant speed of $v_F \in (0, c)$ near the ground. Then the flight time is $T := u_E/v_F$, where u_E denotes the perimeter of the equator.

On the equator, the surface of the earth moves in eastward direction with a speed of $v_E \in (0, c)$, in reference to an inertial system that is at rest at the center of the earth.

Thus by (16.2.4), the velocities of the westbound and of the eastbound jet respectively are (in the inertial system)

$$v_W = \frac{v_F - v_E}{1 - v_F v_E/c^2} \quad , \quad v_O = \frac{v_F + v_E}{1 + v_F v_E/c^2}\,,$$

hence

$$v_W = v_F - v_E + \mathcal{O}\left(\frac{v_{\max}^3}{c^2}\right) \quad , \quad v_O = v_F + v_E + \mathcal{O}\left(\frac{v_{\max}^3}{c^2}\right),$$

where $v_{\max} := \max(v_E, v_F)$.

The time dilations with respect to the inertial system are given by (16.3.3). Thus the difference between the time dilation of the plane and of the earth is on the eastbound flight

$$t_O := T\left(\sqrt{1 - v_O^2/c^2} - \sqrt{1 - v_E^2/c^2}\right) = -u_E \frac{v_E + v_F/2}{c^2} + \mathcal{O}\left(\frac{v_{\max}^3}{c^4}\right).$$

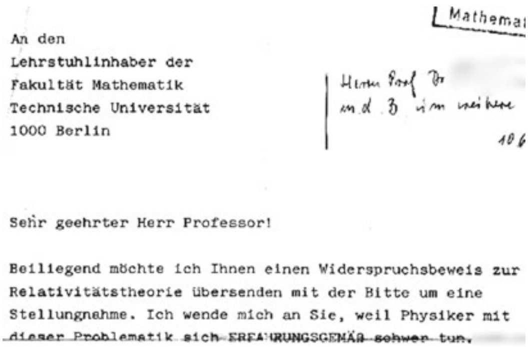
On the westbound flight, it is

$$t_W := T\left(\sqrt{1 - v_W^2/c^2} - \sqrt{1 - v_E^2/c^2}\right) = u_E \frac{v_E - v_F/2}{c^2} + \mathcal{O}\left(\frac{v_{\max}^3}{c^4}\right).$$

Numerically, $v_E = u_E/t_E \approx 465.1\,\mathrm{m/s}$ where $u_E \approx 40\,075\,017\,\mathrm{m}$ is the equatorial perimeter and $t_E \approx 86\,164\,\mathrm{s}$ is the mean length of a sidereal day. If one takes the relative speed of the plane to be $v_F := 900\,\mathrm{km/h} = 250\,\mathrm{m/s}$, one obtains the speeds in the inertial system to be $v_W \approx 215.1\,\mathrm{m/s}$ and $v_O \approx 715.1\,\mathrm{m/s}$. The differences of the time dilations will then be $t_W \approx 152 \times 10^{-9}\mathrm{s}$ and $t_O \approx -263 \times 10^{-9}\mathrm{s}$. ◊

16.14 Exercise (Modified Twin Paradox)

Two (really gutsy) snails travel along the equator, one eastbound, the other westbound. By how much less has the eastbound snail aged compared to the westbound one, when they reconvene at the point of departure? So in Example 16.13, consider the time difference $t_O - t_W$ in the limit $v_F \to 0$ of vanishing speed v_F of the two snails! ◊


An den
Lehrstuhlinhaber der
Fakultät Mathematik
Technische Universität
1000 Berlin

Sehr geehrter Herr Professor!

Beiliegend möchte ich Ihnen einen Widerspruchsbeweis zur Relativitätstheorie übersenden mit der Bitte um eine Stellungnahme. Ich wende mich an Sie, weil Physiker mit dieser Problematik sich ERFAHRUNGSGEMÄß schwer tun.

It is such nonintuitive phenomena that provoke opposition to Einstein's theory of relativity to this very day, and also provoke new attempts at refutation (see the facsimile of such a letter on the right).

16.15 Remark The Minkowski space $\left(\mathbb{R}^4, \langle\cdot,\cdot\rangle_{3,1}\right)$ is used in two roles:

- As *spacetime*. In this case, the points $x \in \mathbb{R}^4$ are called *events* (because events take place at a particular location and a particular time). The (*chronological*) *future* and *past* of an event x are then by definition the open cones

$$I^{\pm}(x) := \left\{y \in \mathbb{R}^4 \mid \langle y - x, y - x\rangle_{3,1} < 0,\ \pm(y_4 - x_4) > 0\right\}.$$

We have already seen that future and past of $0 \in \mathbb{R}^4$ are invariant under orthochronous Lorentz transformations from $\mathrm{L}^{\uparrow}$. More generally, for Poincaré transformations $\Phi_{(a,A)}$, one has

$$\Phi_{(a,A)}\left(I^{\pm}(x)\right) = I^{\pm}\left(\Phi_{(a,A)}(x)\right) \qquad \left(x \in \mathbb{R}^4,\ (a,A) \in \mathbb{R}^4 \times \mathrm{L}^{\uparrow}\right).$$

- As the *tangent space* $T_x\mathbb{R}^4 \cong \mathbb{R}^4$ to spacetime at some point $x \in \mathbb{R}^4$.
  A continuously differentiable curve $c : I \to \mathbb{R}^4$ is called *timelike*, if its tangent vectors $c'(s) \in T_{c(s)}\mathbb{R}^4$ are timelike for all parameters $s \in I$, more specifically *future oriented*, if $c_4'(s) > 0$. Obviously, the future $I^+(x)$ of x consists of the events that can be reached from x by future oriented curves.
  To a curve in space $C : I \to \mathbb{R}^3$ parametrized by 'time' t, one can associate a future oriented curve $c : I \to \mathbb{R}^4,\ t \mapsto \binom{C(t)}{t}$ if the velocity of C satisfies $\|C'(t)\| < 1$ $(t \in I)$.

In *general relativity*, the Minkowski space in its first meaning as spacetime is generalized to a four dimensional manifold with a Lorentz metric.

The translations

$$T_a : \mathbb{R}^4 \to \mathbb{R}^4 \quad , \quad T_a(x) = x + a \qquad \left(a \in \mathbb{R}^4\right)$$

of spacetime $\mathbb{R}^4$, which are contained as a subgroup in the Poincaré group, leave the Lorentz metric invariant. This is called *homogeneity* of spacetime in this context. ◊

16.4 The World from a Relativistic Point of View

"Or high Mathesis, with her charm severe,
Of line and number, was our theme; and we
Sought to behold her unborn progeny,
And thrones reserved in Truth's celestial sphere:
While views, before attained, became more clear;
And how the One of Time, of Space the Three,
Might, in the Chain of Symbols, girdled be"
From the poem *The Tetractys* by W.R. HAMILTON
about the quaternions he found in 1846

In this poem, Hamilton anticipated a fruitful application of his theory. The starting point is the following observation. If we imbed, as in E.27, the space $\mathbb{R}^4$ as the vector space of quaternions,

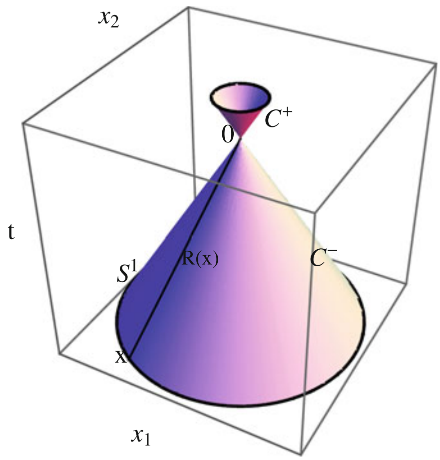
$$\mathcal{I} : \mathbb{R}^4 \to \mathrm{Mat}(2,\mathbb{C}) \quad , \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \mapsto \begin{pmatrix} x_4+ix_3 & x_2+ix_1 \\ -x_2+ix_1 & x_4-ix_3 \end{pmatrix},$$

then the Lorentz metric of the Minkowski space will take the form $\langle x, y\rangle_{3,1} = -\frac{1}{2}\mathrm{tr}\big(\mathcal{I}(x)\mathcal{I}(y)\big)$. The Lorentz transformations of the celestial sphere can be described in this way as well.

This explains the relativistic distortion of the image in the beginning of the chapter. One might conclude from the discussion of the Lorentz contraction (in Example

16.11) that it simply leads optically to a compression of objects along the direction of motion. Images to this effect can be found in older popular expositions of the theory of relativity. But as we will see, this is not the case. In particular, balls still appear as balls to moving observers. An observer in the origin of Minkowski space will see the light[8] from his backwards light cone $\mathcal{C}^-$, and this is independent of whether his velocity is 0 or not. This backwards light cone is the disjoint union of the rays

$$R(x) := \left\{\lambda \left(\begin{smallmatrix} x \\ -1 \end{smallmatrix}\right) \mid \lambda > 0\right\} \quad (x \in S^2)$$



defined by the direction x in space. We call this sphere S^2 the *celestial sphere*.

In the figure on the right, the geometry is shown for two space dimensions (i.e., in $(\mathbb{R}^3, \langle\cdot,\cdot\rangle_{2,1})$), with the celestial sphere S^1.

We now study how the directions and the angles they subtend change under Lorentz transformations. The appropriate language for this purpose consists of mappings called *projectivities*.

16.16 Definition
The **projective linear group** $\mathrm{PGL}(V)$ *of a* $\mathbb{K}$*-vector space* V *is the factor group*

$$\mathrm{PGL}(V) := \mathrm{GL}(V)\,/\,\mathrm{Z}(V)$$

of the general linear group $\mathrm{GL}(V)$ *modulo the normal subgroup*[9] *of dilations*

$$\mathrm{Z}(V) := \{\lambda\,\mathrm{Id}_V \mid \lambda \in \mathbb{K}^*\} \ \triangleleft\ \mathrm{GL}(V)\,.$$

This group acts on the projective space $\mathrm{P}(V)$, because $\mathrm{GL}(V)$ acts on it (see Example E.18), and $\mathrm{Z}(V)$ leaves the 1-dimensional subspaces of V invariant.

16.17 Example (Projective Space $\mathbb{K}\mathrm{P}(1)$ and Möbius Transformations)
As shown in Example 6.52, the real projective space $\mathbb{R}\mathrm{P}(1) \equiv \mathrm{P}(\mathbb{R}^2)$ is diffeomorphic to the circle S^1. Analogously, the complex projective space $\mathbb{C}\mathrm{P}(1) \equiv \mathrm{P}(\mathbb{C}^2)$ is diffeomorphic to the sphere S^2 (Remark 6.36).

Another way to see this is by identifying sphere and projective space with $\mathbb{K}\cup\{\infty\}$, for $\mathbb{K} = \mathbb{R}$ and $\mathbb{C}$ respectively.

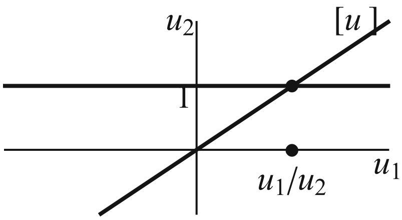
[8] And more generally electromagnetic radiation to the extent that it doesn't propagate in a medium, with speed below the speed of light 1.

[9] $\mathrm{Z}(V)$ is also called the *center* of $\mathrm{GL}(V)$, because it consists of those mappings that commute with all elements of the group.

- In the case of S^1 and S^2 respectively, this is done by stereographic projection (Example A.29.3).
- In the case of $\mathbb{K}\mathrm{P}(1)$, one identifies the equivalence class

$$[v] = \mathrm{span}(u) \setminus \{0\} \ \in \mathbb{K}\mathrm{P}(1)$$

of $u \in \mathbb{K}^2 \setminus \{0\}$ with ∞ if $u_2 = 0$, and otherwise with $u_1/u_2 \in \mathbb{K}$, for $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$ respectively. Geometrically, the latter corresponds to the intersection at $\{(u_1/u_2, 1)\} = [u] \cap \big(\mathbb{K} \times \{1\}\big)$, see the figure.



In this notation, the *projectivities*, or *Möbius transformations* for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}(2, \mathbb{K})$ are of the form

$$z \mapsto \frac{az+b}{cz+d} \ \text{ for } z = \frac{u_1}{u_2}\,, \quad \text{and} \quad z \mapsto \frac{a}{c} \ \text{ for } z = \infty\,,$$

see also Exercise 6.42. They only depend on three parameters from $\mathbb{K}$, because numerator and denominator may be multiplied with the same number from $\mathbb{K}^*$ without changing the result.

Thus $\mathrm{PGL}(\mathbb{R}^2)$ is a 3-dimensional, and $\mathrm{PGL}(\mathbb{C}^2)$ a 6-dimensional Lie group.

In the article [AR] by ARNOLD and ROGNESS, it is shown and visualized that the Möbius transformations from $\mathrm{PGL}(\mathbb{R}^2)$ arise by composition of the (inverse) stereographic mapping and rigid motion of the sphere S^2 in $\mathbb{R}^3$. ◊

The following theorem was proved independently by PENROSE in [Pen] and by TERRELL in [Te] in 1959.

16.18 Theorem *The restricted Lorentz group* $\mathrm{SO}^+(3,1)$ *acts on the celestial sphere* S^2 *as the group* $\mathrm{PSL}(2,\mathbb{C}) := \mathrm{SL}(2,\mathbb{C})/\{\pm \mathbb{1}\}$ *of orientation preserving Möbius transformations.*

Proof:
- Given the linear isomorphism

$$A : \mathbb{R}^4 \to \mathrm{Sym}(2,\mathbb{C})\,,\ A(v) := \left(\begin{smallmatrix} v_4+v_3 & v_1+\mathrm{i}v_2 \\ v_1-\mathrm{i}v_2 & v_4-v_3 \end{smallmatrix}\right) \quad \text{with } \det\big(A(v)\big) = -\langle v, v\rangle_{3,1}\,,$$

let us first study the group action

$$\Phi : \mathrm{SL}(2,\mathbb{C}) \times \mathbb{R}^4 \to \mathbb{R}^4 \quad , \quad \Phi_g(v) := A^{-1}\big(gA(v)g^*\big)\,. \tag{16.4.1}$$

The Lorentz metric $w := \Phi_g(v)$ satisfies

$$\langle w, w\rangle_{3,1} = -\det\big(gA(v)g^*\big) = -|\det(g)|^2 \det\big(A(v)\big) = -\det\big(A(v)\big) = \langle v, v\rangle_{3,1} \ .$$

Therefore, the group $\mathrm{SL}(2, \mathbb{C})$ acts by Lorentz transformations, and we obtain a group homomorphism

$$\tilde{\Pi} : \mathrm{SL}(2, \mathbb{C}) \to \mathrm{O}(3, 1) \, .$$

• This is an extension of the group homomorphism $\Pi : \mathrm{SU}(2) \to \mathrm{SO}(3)$ from Theorem E.29. Indeed, $\mathrm{SU}(2) \le \mathrm{SL}(2, \mathbb{C})$ and $\mathrm{SO}(3) \le \mathrm{O}(3, 1)$ are subgroups, and for $g \in \mathrm{SU}(2)$, one has $g^* = g^{-1}$. Furthermore, in terms of the mapping $\sigma : \mathbb{R}^3 \to \mathfrak{su}(2)$, $x \mapsto \frac{1}{2}\left(\begin{smallmatrix} -\mathrm{i}x_3 & -\mathrm{i}x_1+x_2 \\ -\mathrm{i}x_1-x_2 & \mathrm{i}x_3 \end{smallmatrix}\right)$, one has

$$A\big(\left(\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right)\big) = 2\mathrm{i}\,\sigma(x) \quad \text{, thus} \quad \Phi_U\big(\left(\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right)\big) = \left(\begin{smallmatrix} \Pi_U(x) \\ 0 \end{smallmatrix}\right) \quad \big(x \in \mathbb{R}^3,\ U \in \mathrm{SU}(2)\big).$$

• The image $\tilde{\Pi}\big(\mathrm{SL}(2, \mathbb{C})\big)$ is contained in the *restricted* Lorentz group $\mathrm{SO}^+(3, 1)$:

– by Remark 16.4.2, the subgroup $\mathrm{SO}^+(3, 1) \le \mathrm{O}(3, 1)$ is the connected component of the identity;
– the matrices $g \in \mathrm{SL}(2, \mathbb{C})$ have a polar decomposition $g = u \exp(w)$ with $u \in \mathrm{SU}(2)$ and $w \in \mathrm{Mat}(2, \mathbb{C})$ Hermitian and traceless. Conversely, every such product $u \exp(w)$ lies in $\mathrm{SL}(2, \mathbb{C})$. Since $\mathrm{SU}(2) \cong S^3$, it follows that $\mathrm{SL}(2, \mathbb{C}) \cong S^3 \times \mathbb{R}^3$ is connected.

• On the other hand, the image of $\mathrm{SL}(2, \mathbb{C})$ *is equal to* $\mathrm{SO}^+(3, 1)$, as can be seen from $\tilde{\Pi}\big(\mathrm{SU}(2)\big) = \mathrm{SO}(3)$ and the transformation of the positive factor in the polar decomposition of $g \in \mathrm{SL}(2, \mathbb{C})$.
• We now note that the mapping

$$\mathbb{C}^2 \to \mathrm{Sym}(2, \mathbb{C}) \quad , \quad z = \left(\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right) \mapsto z \otimes \overline{z}^\top = \begin{pmatrix} |z_1|^2 & z_1\overline{z_2} \\ z_2\overline{z_1} & |z_2|^2 \end{pmatrix}$$

can be composed with $A^{-1} : \mathrm{Sym}(2, \mathbb{C}) \to \mathbb{R}^4$ to a mapping that is called the *Hopf map*

$$\mathrm{Hopf} : \mathbb{C}^2 \to \mathbb{R}^4 \quad , \quad z = \left(\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right) \mapsto \begin{pmatrix} 2\mathrm{Re}(z_1\overline{z_2}) \\ 2\mathrm{Im}(z_1\overline{z_2}) \\ |z_1|^2-|z_2|^2 \\ |z_1|^2+|z_2|^2 \end{pmatrix} .$$

The first three (namely the 'spatial') components were already used in connection with the 2-dimensional harmonic oscillator for the parametrization of its orbit space S^2, see Remark 6.36.

The image $\mathrm{Hopf}\big(\mathbb{C}^2 \setminus \{0\}\big) \subset \mathbb{R}^4$ is the forward light cone $\mathcal{C}^+$ from Definition 16.7, and for $v := \mathrm{Hopf}(z) \in \mathcal{C}^+$, the fiber is the circle

$$\mathrm{Hopf}^{-1}(v) = \{\lambda z \mid \lambda \in S^1 \subset \mathbb{C}\} \, .$$

The linear action $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mapsto \begin{pmatrix} g_{1,1}z_1+g_{1,2}z_2 \\ g_{2,1}z_1+g_{2,2}z_2 \end{pmatrix}$ of the matrix $g = \begin{pmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C})$ on $\mathbb{C}^2$ thus leads, by composition with the Hopf map, to a Lorentz transformation of the forward light cone, which transforms the celestial sphere conformally. □

Some Möbius transformations of the celestial sphere are shown in Figures 16.4.1 and 16.4.2; more precisely, their generating vector fields are shown:

Figure 16.4.1 Left: rotation: Middle: boost; Right: combination of rotation and boost

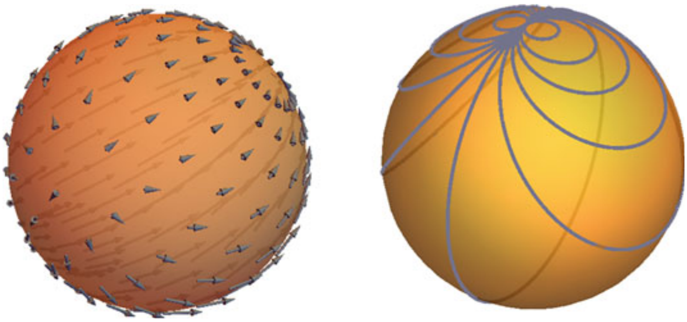


Figure 16.4.2 Left: Combination of rotation and boost, with different axes; Right: Orbits of a degenerate Möbius vector field

- The daily apparent *rotation* of the stars about the pole-to-pole axis of the earth, which is of course familiar.
- The *boost* concentrates the stars in the direction of the velocity, whereas their density is increased in the inverse direction of flight.
- *Combinations* of rotation and boost are the more typical case. It is only for rotations about the velocity vector that the two zeros of the Möbius vector field are antipodes.
- The two zeros can merge into a single *degenerate zero*.

Since the (angle preserving) Möbius transformations map circles into circles, no relativistic length contraction is visible. This observation is no contradiction to the previous remarks, because in the Lorentz contraction studied in Example 16.11, the two ends of the rod that was passing by were measured at different locations in the reference frame. In Penrose and Rindler [PR1], one can find a more extensive discussion.

16.5 From Einstein to Galilei—and Back

Galilei's theory of relativity is obtained from Einstein's special theory of relativity by a limiting process, where the speed of light c goes to infinity. To see this, we insert $c > 0$ into the Minkowski product[10]:

$$\langle\cdot,\cdot\rangle_c : \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}, \quad , \quad \langle v, w\rangle_c = v_1 w_1 + v_2 w_2 + v_3 w_3 - c^2 v_4 w_4 \,, \tag{16.5.1}$$

and we define the Lorentz groups as the invariance groups of this Minkowski product, i.e.,

$$\mathrm{L}_c := \left\{ A \in \mathrm{GL}(4,\mathbb{R}) \mid \forall v, w \in \mathbb{R}^4 : \langle Av, Aw\rangle_c = \langle v, w\rangle_c \right\}.$$

In the polar decomposition $A = \tilde{O} L_c(v)$ of a matrix $A \in \mathrm{L}_c$ with an orthogonal matrix $\tilde{O}$, we obtain, by modification of (16.2.5), the explicit form of the Lorentz boost:

$$L_c(v) = \gamma_c(v) \begin{pmatrix} (\mathbb{1}_3 - P_v)/\gamma_c(v) + P_v & v \\ v^\top/c^2 & 1 \end{pmatrix} \qquad \left(v \in \mathbb{R}^3,\ 0 < \|v\| < c\right), \tag{16.5.2}$$

with $\gamma_c(v) := (1 - \|v\|^2/c^2)^{-1/2} = 1 + \mathcal{O}(c^{-2})$. Thus for every velocity $v \in \mathbb{R}^3$, one has

$$L_\infty(v) := \lim_{c\to\infty} L_c(v) = \begin{pmatrix} \mathbb{1}_3 & v \\ 0 & 1 \end{pmatrix}.$$

If A belongs to the restricted Lorentz group, then the orthogonal matrix in the polar decomposition has the form $\tilde{O} = \left(\begin{smallmatrix} O & 0 \\ 0 & 1 \end{smallmatrix}\right)$ with a rotation matrix $O \in \mathrm{SO}(3)$. An even simpler form than $\tilde{O} L_\infty(v) = \left(\begin{smallmatrix} O & Ov \\ 0 & 1 \end{smallmatrix}\right)$ is obtained if we exchange the factors in the polar decomposition, because $L_\infty(v)\tilde{O} = \left(\begin{smallmatrix} O & v \\ 0 & 1 \end{smallmatrix}\right)$.

The Poincaré group in its explicitly c-dependent form is the semidirect product $P_c := \mathbb{R}^4 \rtimes \mathrm{L}_c$. It can be written as a matrix group by combining $a \in \mathbb{R}^4$ and $A \in \mathrm{L}_c$ into the 5×5-matrix $\left(\begin{smallmatrix} A & a \\ 0 & 1 \end{smallmatrix}\right)$. After taking the limit $c \to \infty$ and writing the translation vector a as $a := \left(\begin{smallmatrix} q \\ t \end{smallmatrix}\right) \in \mathbb{R}^4 = \mathbb{R}^3 \times \mathbb{R}$, we obtain the elements $\left(\begin{smallmatrix} O & v & q \\ 0 & 1 & t \\ 0 & 0 & 1 \end{smallmatrix}\right)$ of the (proper

[10] It can easily be determined by a consideration of physical dimensions (units) where the factors c for conversion between space and time need to be inserted.

orthochronous) *Galilei group*, with the multiplication law

$$\begin{pmatrix} O_1 & v_1 & q_1 \\ 0 & 1 & t_1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} O_2 & v_2 & q_2 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} O_1 O_2 & v_1 + O_1 v_2 & q_1 + v_1 t_2 + O_1 q_2 \\ 0 & 1 & t_1 + t_2 \\ 0 & 0 & 1 \end{pmatrix}. \tag{16.5.3}$$

Just as the Poincaré group, this one is also a 10-dimensional Lie group.

16.19 Remarks (Galilei group)

1. Thus this group is isomorphic to the semidirect product $\mathbb{R}^4 \rtimes \mathbb{E}(3)$ of the Euclidean group $\mathbb{E}(3)$ with $\mathbb{R}^4$: The group element $g := (\Delta q, \Delta t\,;\ v, O) \in \mathbb{R}^4 \rtimes \mathbb{E}(3)$ acts on the spacetime point $(q, t) \in \mathbb{R}^4$ by
$$\Phi_g(q, t) := \big(Oq + vt + \Delta q,\ t + \Delta t\big). \tag{16.5.4}$$
2. The Minkowski bilinear form (16.5.1) itself does not have a limit for infinite speed of light.
Instead, in Galilei's theory of relativity, there exists an absolute time that is independent of the reference frame. Two points $(q, t) \neq (q', t')$ in spacetime are called *simultaneous* if there is no physical way of reaching (q', t') from (q, t), i.e., if there is no $(v, s) \in \mathbb{R}^4$ such that $(q', t') = (q + vs, t + s)$. This is obviously the case if and only if $t = t'$. Thus Galilei-simultaneity defines[11] an equivalence relation on spacetime.
Consequently, spacetime can be viewed as a bundle over the time axis $\mathbb{R}$, whose fibers are isomorphic (but not canonically isomorphic) to a 3-dimensional affine space. This aspect is discussed further for instance in Chapter II.2 of the book [Scho] by SCHOTTENLOHER.
This global time can only be changed by an explicit shift in time, but not by a change in velocity. This can be seen from the entry $t_1 + t_2$ in the law of multiplication (16.5.3).
The point in space however depends on the velocity vector also in Galilei's theory of relativity, in accordance with the entry $q_1 + v_1 t_2 + O_1 q_2$ in (16.5.3). So there is no absolute notion of 'equal location'. To also abandon the notion of an absolute simultaneity amounted to the crucial step toward the special theory of relativity.
3. More amazing than the fact that the Galilei group can be obtained by a limiting process from the Poincaré group is the fact that the reverse route is also possible. By means of a technique called *deformation*, Einstein's special theory of relativity can be derived group theoretically from Galilei's theory with almost no ingredients from physics.
The Lie algebra $\mathfrak{g}$ of a Lie group determines the local structure of the Lie group (see Remark E.23.2). In a basis $b_1, \ldots, b_n$ of the $\mathbb{R}$-vector space $\mathfrak{g}$, the Lie algebra structure is determined by the coefficients $c_{i,j}^k \in \mathbb{R}$ in their Lie bracket $[b_i, b_j] = \sum_{k=1}^n c_{i,j}^k b_k$. The antisymmetry of the Lie bracket implies $c_{j,i}^k = -c_{i,j}^k$, whereas the Jacobi identity yields the quadratic relations

[11] In contrast to the special theory of relativity!

$\sum_{\ell=1}^{n}\left(c_{i,j}^{\ell}c_{\ell,k}^{m}+c_{j,k}^{\ell}c_{\ell,i}^{m}+c_{k,i}^{\ell}c_{\ell,j}^{m}\right)=0$ for the coefficients.
Thus the set of structure tensors $(c_{i,j}^{k})$ of Lie algebras is a subset $\mathcal{L}(n)$ of an n^3-dimensional $\mathbb{R}$-vector space. The group $\mathrm{GL}(n,\mathbb{R})$ acts on this vector space, and its orbits consist of mutually isomorphic Lie algebras.
The *contractions* of a Lie algebra correspond to the boundary points of its orbits. As we have seen, the Lie algebra of the Galilei group is obtained by such a contraction from the Lie algebra of the Poincaré group.
The basic idea of *deformation* of $\mathfrak{g}$ consists conversely of perturbing the coefficients of $\mathfrak{g}$ in $\mathcal{L}(n)$ and looking whether the Lie algebra $\mathfrak{g}'$ thus obtained is isomorphic to $\mathfrak{g}$ (see Chap. 7.2 of the book [OV] by ONISHCHIK and VINBERG). For the Lie algebra of the Euclidean group $\mathbb{E}(3)$, one is led by this process to the Lie algebra $\mathfrak{so}(4)$ of the rotation group on one hand, and to the one of the Lorentz group on the other hand. The rotation group can be dismissed as a candidate for a relativistic symmetry of spacetime. Details can be found in [FOF] and work quoted there. ◊

The fundamental interactions of a theory in physics should be invariant under the relativistic symmetry transformations. This limits the form of the physically fundamental Hamilton function.

In order to define this invariance formally, we need to consider phase spaces that are cotangent bundles over spacetime, rather than such over space only.

Quite generally, we begin with a possibly time dependent Hamiltonian $H : T^*M \times \mathbb{R}_t \to \mathbb{R}$ on the extended phase space T^*M (namely extended by the time axis $\mathbb{R}_t$), with a configuration manifold M. The product $T^*M \times T^*\mathbb{R}_t$ of the cotangent spaces (T^*M, ω_1) and $(T^*\mathbb{R}_t, \omega_2)$ with their canonical symplectic forms has the symplectic structure given by Theorem 6.48,

$$\omega := \omega_1 \ominus \omega_2 = \pi_1^*(\omega_1) - \pi_2^*(\omega_2)\,, \tag{16.5.5}$$

where the π_i are the projections on the factors. The sign is chosen in such a way that ω is adapted to the Minkowski space $\langle\cdot,\cdot\rangle_{3,1}$.

We compare the dynamics generated by H with the one generated by

$$\tilde{H} : T^*(M \times \mathbb{R}_t) \to \mathbb{R} \quad , \quad \tilde{H}(p, E; q, t) := H(p, q, t) - E\,. \tag{16.5.6}$$

The Hamiltonian equations of motion, written in local canonical coordinates $p = (p_1, \ldots, p_d)$, $q = (q_1, \ldots, q_d)$ of T^*M, are

$$\dot{p}_j = -\frac{\partial H}{\partial q_j} \quad , \quad \dot{q}_j = \frac{\partial H}{\partial p_j} \quad , \quad \dot{E} = \frac{\partial H}{\partial t} \quad \text{and} \quad \dot{t} = 1\,,$$

where the dot denotes the derivative with respect to the time parameter s. Thus we can set $t = s$, and the solutions for $\tilde{H}$ correspond to those for H.

Moreover, one has $E = H(p, q, t)$ on the level set $\tilde{H}^{-1}(0)$, so the phase space variable E can be interpreted as total energy. We now specialize to the case $M := \mathbb{R}_q^3$, and study the form of those Hamiltonians that are invariant under the action of the Poincaré group $P = \mathbb{R}^4 \rtimes \mathrm{L}$, lifted to $T^*\mathbb{R}^4$. Here $(a, A) \in P$ acts by the affine map $\Phi_{(a,A)} : \mathbb{R}^4 \to \mathbb{R}^4,\ x \mapsto Ax+a$ on spacetime $\mathbb{R}^4$, and therefore by the ω-symplectic cotangent lift from Definition 10.32,

$$\Phi^{T^*}_{(a,A)} : T^*\mathbb{R}^4 \to T^*\mathbb{R}^4 \quad , \quad (p, x) \mapsto \big(I\big(A^{-1}\big)^\top I\, p,\ Ax + a\big) \tag{16.5.7}$$

on its cotangent space. Due to the relation $A^\top I A = I$, with the diagonal matrix $I = \mathrm{diag}(1, 1, 1, -1)$, which follows from the definition of the Lorentz group according to Remark 16.2.4, we can simply write A instead of the matrix $I\big(A^{-1}\big)^\top I$.

16.20 Lemma (Cotangent Lift of the Galilei Transformation)
The nonrelativistic limit of (16.5.7) for the Poincaré transformation with polar decomposition $A_c := L_c(v)\tilde{O}$, *orthogonal matrix* $\tilde{O} = \left(\begin{smallmatrix} O & 0 \\ 0 & 1 \end{smallmatrix}\right)$, *and* $a = \left(\begin{smallmatrix} \Delta q \\ \Delta t \end{smallmatrix}\right)$ *is*

$$(p, E\,;\, q, t) \longmapsto \big(Op\,,\ E + \langle v, Op\rangle\ ;\ Oq + vt + \Delta q\,,\ t + \Delta t\big)\,. \tag{16.5.8}$$

This is the cotangent lift of the Galilei transformation (16.5.4) on $T^*\mathbb{R}^4 \cong T^*\mathbb{R}^3 \times T^*\mathbb{R}$.

Proof: Analogous to the above remark, one has $I\big(A_c^{-1}\big)^\top I = D_c A_c D_c^{-1}$, with $D_c := \mathrm{diag}(1, 1, 1, c^2)$. Moreover,

$$D_c A_c D_c^{-1} = D_c L_c(v) D_c^{-1}\ D_c \tilde{O} D_c^{-1} = D_c L_c(v) D_c^{-1}\ \tilde{O}\,.$$

The limit $c \to \infty$ is obtained after conjugating formula (16.5.2) for the Lorentz boost $L_c(v)$. □

The remarkable thing about (16.5.8) is that under this limit, the momentum $p \in \mathbb{R}^3$ does not change when transitioning to the reference frame with relative velocity v. But we need to remember that it is only by giving a Hamilton function for a particle that the momentum acquires a connection to the velocity of that particle.

As $a \in \mathbb{R}^4$ may be chosen freely, a Lorentz invariant Hamilton function can only depend on the 4-momentum p. For arbitrary $f \in C^1(\mathbb{R}, \mathbb{R})$, a Hamilton function of the form

$$\tilde{H}_c : T^*\mathbb{R}^4 \to \mathbb{R} \quad , \quad \tilde{H}_c(p, q) := f\big(\langle p, p\rangle_{1/c}\big)$$

is invariant. In the simplest case, $f : \mathbb{R} \to \mathbb{R}$ is linear, and in comparison with the nonrelativistic theory, we denote the slope of f as $1/(2m)$. Writing the 4-momentum in the form $(p, E) \in \mathbb{R}^3 \times \mathbb{R}$, one obtains for the value $-\frac{mc^2}{2}$ of the Hamilton function

$$\tilde{H}_c(p, E; q, t) := \tfrac{\|p\|^2}{2m} - \tfrac{E^2}{2mc^2}$$

the formula $E = mc^2\sqrt{1+\left(\frac{\|p\|}{mc}\right)^2} = mc^2 + \frac{\|p\|^2}{2m} + \mathcal{O}\left(\left(\frac{\|p\|}{mc}\right)^4\right)$ for the energy of the relativistic free particle. On the level set of $\frac{mc^2}{2}$, the equations of motion

$$\frac{\mathrm{d}p}{\mathrm{d}s} = 0 \quad , \quad \frac{\mathrm{d}E}{\mathrm{d}s} = 0 \quad , \quad \frac{\mathrm{d}q}{\mathrm{d}s} = \frac{p}{m} \quad , \quad \frac{\mathrm{d}t}{\mathrm{d}s} = \frac{E}{mc^2}$$

of $\tilde{H}$ lead to the relativistic relation

$$\boxed{\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{p}{m\sqrt{1+\left(\frac{\|p\|}{mc}\right)^2}}}$$

between velocity and momentum.

Whereas the pointwise limit $\lim_{c\to\infty} \tilde{H}_c(p, E; q, t)$ equals $\frac{\|p\|^2}{2m}$, for given values of p (as well as q and t), on the c-dependent level set of $\frac{mc^2}{2}$, the nonrelativistic limit of $\tilde{H}_c(p, \tilde{E} + mc^2; q, t)$ will be $\frac{\|p\|^2}{2m} - \tilde{E}$.

But the nonrelativistic Hamilton function

$$\tilde{H} : T^*(\mathbb{R}^3_q \times \mathbb{R}_t) \to \mathbb{R} \quad , \quad \tilde{H}(p, E; q, t) := \frac{\|p\|^2}{2m} - E \tag{16.5.9}$$

obtained in this way, or by transition (16.5.6) to the extended phase space, is not invariant under the lifted action (16.5.8) of the Galilei group.

This is why, rather than (16.5.8), one uses the following transformations for the nonrelativistic limit:

16.21 Lemma (Phase Space Action of the Galilei Group) *The mappings*

$$(p, E\,;\, q, t) \mapsto \left(Op + mv\,,\; E + \langle v, Op\rangle + \tfrac{1}{2}m\|v\|^2\,;\; Oq + vt + \Delta q\,,\; t + \Delta t\right) \tag{16.5.10}$$

define an ω-symplectic group action (with respect to ω from 16.5.5) of the Galilei group on the phase space $T^*\mathbb{R}^4$ of spacetime, and this action covers the action Φ from Remark 16.19.1.

Proof:

- The mapping (16.5.10) is the composition of the (ω-symplectic) cotangent lift (16.5.8) with the translation of the fibers of $T^*\mathbb{R}^4$ by the constant $\left(mv, \frac{1}{2}m\|v\|^2\,;\; 0, 0\right)$. It is therefore ω-symplectic.

• Moreover, the composition of the actions of the motions $(v_1, O_1), (v_2, O_2) \in \mathbb{E}(3)$ satisfies:

$$O_1(O_2 p + m v_2) + m v_1 = O_1 O_2 p + m(O_1 v_2 + v_1)$$

and

$$\begin{aligned}\left(E + \langle v_2, O_2 p\rangle + \tfrac{1}{2} m \|v_2\|^2\right) + \langle v_1, O_1(O_2 p + m v_2)\rangle + \tfrac{1}{2} m \|v_1\|^2 \\ = E + \langle O_1 v_2 + v_1, O_1 O_2 p\rangle + \tfrac{1}{2} m \|O_1 v_2 + v_1\|^2 .\end{aligned}$$

So we indeed have a group action by the Galilei group.
• The last two entries in (16.5.10) coincide with (16.5.4). So the action Φ is indeed covered. □

While these mappings differ from (16.5.8) merely by a constant fiber translation of $T^*\mathbb{R}^4$, they do leave the Hamilton function (16.5.9) invariant.

Nonrelativistic particles can interact in many Galilei-invariant manners. The extended phase space of n particles is

$$P_n := T^*\left(\mathbb{R}_q^{3n} \times \mathbb{R}_t\right) .$$

The action of the Galilei group is diagonal with respect to the coordinates of the particles, i.e., for particle masses $m_k > 0$ and the total mass $m := \sum_{k=1}^n m_k$, the point $(p_1, \ldots, p_n, E; q_1, \ldots, q_n, t) \in P_n$ in phase space is transformed into

$$\begin{aligned}\big(O(p_1 + m_1 v), \ldots, O(p_n + m_n v)\,,\; E + \langle v, p_1 + \ldots + p_n\rangle + \tfrac{1}{2} m \|v\|^2\,; \\ O(q_1 + vt) + \Delta q, \ldots, O(q_n + vt) + \Delta q\,,\; t + \Delta t\big) .\end{aligned}$$

16.22 Exercise (Galilei Group)

(a) Show in analogy to Lemma 16.21 that these mappings define a symplectic action of the Galilei group on the extended phase space P_n of n particles.
(b) We consider the Hamiltonian $H : P_n \to \mathbb{R}$,

$$H(p_1, \ldots, p_n, E; q_1, \ldots, q_n, t) = \sum_{k=1}^n \frac{\|p_k\|^2}{2m_k} + \sum_{1 \le k < \ell \le n} V_{k,\ell}(q_k - q_\ell) - E$$

with the $V_{k,\ell} \in C^\infty(\mathbb{R}^3, \mathbb{R})$. Show that H is Galilei invariant, if the interaction potentials $V_{k,\ell}$ are rotation invariant. ◊

16.23 Remark (Structure of Relativistic Theories)
In contrast, a finite number of *relativistic* particles cannot interact in a Poincaré invariant manner. This was first shown with regard to two particles (see CURRIE, JORDAN and SUDARSHAN in [CJS]), and then generalized by LEUTWYLER in [Leu]. In [AB], ARENS and BABBITT treat interactions that do not need to be Hamiltonian.

These results indicate that a relativistic theory with interactions needs to admit infinitely many particles. ◊

16.6 Relativistic Dynamics

Two non-parallel affine hyperplanes of $\mathbb{R}^4$ intersect. Thus for accelerated frames of reference, there exist points in spacetime that are simultaneous to *different* points on the worldline, see Figure 16.6.1, left.

16.24 Example (Constant Force)
In the case of constant force (and thus constant acceleration $g > 0$ in the reference frame that moves along with the particle), the orbit is described by the Hamiltonian

$$H : \mathbb{R}^2 \to \mathbb{R} \quad , \quad H(p, q) = \sqrt{1 + p^2} - g\,q\,;$$

compare this with Example 8.7. This yields $\dot{p} = g$ and $\dot{q} = p/\sqrt{1 + p^2}$.

All solution curves are obtained by a spacetime translation from the one with energy $H = 0$ and $p(0) = 0$, thus $q(0) = 1/g$. We obtain $\dot{q} = \sqrt{1 - g/q^2}$, and therefore

$$q(t) = \sqrt{g^{-2} + t^2} \quad \text{and} \quad \dot{q}(t) = \frac{t}{\sqrt{g^{-2} + t^2}} \in (-1, 1)\,.$$

The worldline $t \mapsto \big(q(t), t\big) \in \mathbb{R}^2$ is the parametric equation of a hyperbola (see Figure 16.6.1, right).

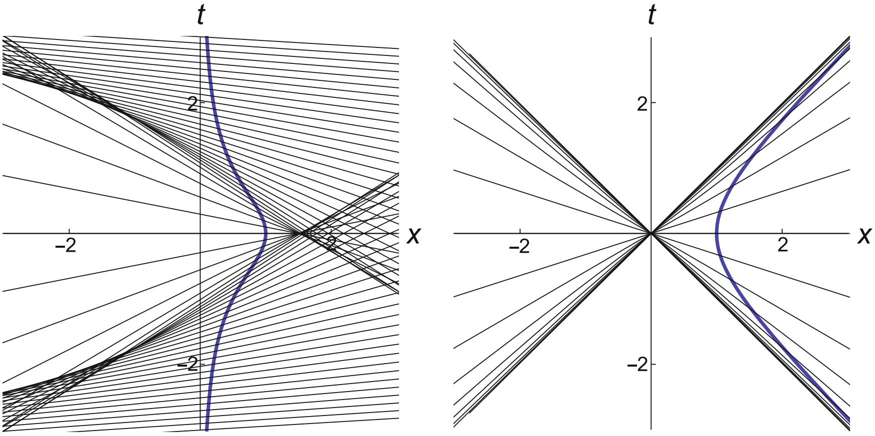


Figure 16.6.1 Worldlines, i.e., trajectories in the spacetime representation. The straight lines are simultaneous in the accelerated frame of reference. Left: Return to the starting point. Right: Constant acceleration

The straight lines in spacetime consisting of those events that are simultaneous to the point $\big(q(s), s\big)$ in the frame of reference that moves along are therefore of the form

$$\big\{\big(q(s), s\big) + c\big(1, \dot{q}(s)\big) \mid c \in \mathbb{R}\big\} = \operatorname{span}\big(\sqrt{g^{-2} + s^2}, s\big) \qquad (s \in \mathbb{R}),$$

so they all intersect in the origin of spacetime.

Hence the double cone $\{(x, t) \in \mathbb{R}^2 \mid t^2 > x^2\}$ of those events that are timelike with respect to the origin of the Minkowski space $\mathbb{R}^2$ does not contain any point that is simultaneous to a point of the worldline. ♢

The phenomenon described here does not match our everyday experience, but it also rarely occurs in this experience: accelerations much larger than the acceleration of gravity and distances larger than a light year are not part of our experience.

16.25 Exercise (Constant Acceleration)
Calculate the distance of the point of intersection in Example 16.24 from an accelerated observer, if said observer experiences a constant acceleration of $10\,\mathrm{m/s^2}$. ♢

Chapter 17
Symplectic Topology

Volume preserving, non-symplectic camel. Image: courtesy of Norbert Nacke

In the theory of dynamical systems, topological methods are often employed when the dynamics is too complicated to answer questions like the one about the existence of periodic orbits directly.

As Hamiltonian differential equations (and also gradient systems) are defined in terms of the derivative of a single function $H : M \to \mathbb{R}$, topological statements about the zeros of a real function on a manifold turn into statements about dynamics. For instance, on a compact manifold, such a function attains a minimum and a maximum.

Morse theory predicts (depending on the topology of the manifold) the existence of further critical points of the Hamilton function. All these critical points are equilibria for the dynamical system.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7_17

Most phase spaces in classical mechanics are not compact, and therefore the above arguments need to be refined. Symplectic topology, a very active research area during the past decades, attempts to fathom such dynamical properties of Hamiltonian systems.

But this is not an exhaustive summary of its scope. For instance, by the theorem of Darboux (see page 229), symplectic manifolds of the same dimension cannot be distinguished *locally*, in stark contrast to Riemannian manifolds. But what are their *global* invariants? Some answers come to mind immediately:

- There can exist non-isomorphic symplectic structures on a manifold, since for instance the volume of a compact symplectic manifold is an invariant.
- Also, not every manifold carries a symplectic structure. Apart from the condition of even dimension, the manifold needs to be orientable.

In this chapter, we will address a few more extensive answers.

17.1 The Symplectic Camel and the Eye of a Needle

"It is easier for a camel to go through the eye of a needle, than for a rich man to enter into the kingdom of God." Gospel of Mark, 10:25 (KJV)

This is the curious title under which an analysis of invariants of symplectic mappings, started by M. Gromov in the 1980s, is given. The idea is: If the symplectic camel could be deformed under arbitrary volume preserving diffeomorphisms, rather than only under symplectic ones, then it would not be difficult for the camel to pass through the eye of the needle. As it stands however, a symplectic ribcage denies him the passage.

The camel is modeled by a ball B_r of radius r in a symplectic vector space $(\mathbb{R}^{2n}, \omega_0)$, and the eye of the needle by a hole of radius R in a hypersurface $H \subset \mathbb{R}^{2n}$. For $n = 1$ degrees of freedom, the camel can always form a thin neck to slip through the hole. For $n \geq 2$ however, it is only for $r < R$ that this is symplectically possible.[1]

In the *nonsqueezing* theorem proved by Gromov in 1985, one asks when the ball B_r can be mapped symplectically into a special cylinder Z_R of radius R. This is possible in an arbitrary dimension if and only if $r \leq R$. The nonsqueezing theorem implies the above statement about the symplectic camel, but also a lot more, as we shall see (confer however ABBONDANDOLO, MATVEYEV [AbMa]).

We first study the problem in a radically simplified situation, in which we only consider *affine symplectic mappings* of the $2n$-dimensional vector space E:

$$f : E \to E \quad , \quad x \mapsto g(x) + a \qquad \text{with } g \in \mathrm{Sp}(E, \omega) \text{ and } a \in E\,. \tag{17.1.1}$$

[1] in technical terms: There exists a symplectic isotopy $\Phi_t : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $t \in [0, 1]$ with $\Phi_0 = \mathrm{Id}$ that leaves the punctured hypersurface invariant and for which $\Phi_1(B_r)$ lies in the opposite component of $\mathbb{R}^{2n} \setminus H$ from the one containing B_r.

These mappings form a semidirect product $E \rtimes \mathrm{Sp}(E, \omega)$, and a Lie group, called $\mathrm{ASp}(E)$. [2]

In order to define objects like balls or cylinders, one actually needs an additional structure, namely a Euclidean norm. Moreover, balls are not mapped into balls under affine symplectic mappings (in other words, the property of being a ball is not an affine symplectic invariant).

In contrast, in a finite dimensional real vector space E, we can define without a norm what it means for a subset $\mathcal{E} \subseteq E$ to be an *ellipsoid*, namely it means that there exists an appropriate positive definite quadratic form

$$Q : E \to \mathbb{R} \qquad \text{for which } \mathcal{E} = \{x \in E \mid Q(x) \le 1\} .$$

As the quadratic form can be reconstructed from the ellipsoid by

$$Q(0) := 0 \quad \text{and} \quad Q(x) := \inf\{q > 0 \mid x/q \in \mathcal{E}\} \text{ for } x \in E \setminus \{0\} ,$$

ellipsoids and positive definite quadratic forms are two aspects of the same thing. Our first question is about symplectic normal forms of ellipsoids (respectively, positive definite quadratic forms). We first draw a comparison with normal forms under other groups:

- On the space $\mathcal{P} = \mathcal{P}(E)$ of such forms, the general linear group $\mathrm{GL}(E)$ acts transitively by
$$\mathrm{GL}(E) \times \mathcal{P} \to \mathcal{P} \quad , \quad (f, Q) \mapsto Q \circ f .$$
This can be seen for the case of $E = \mathbb{R}^d$ and the representation $Q(x) = x^\top \mathcal{Q} x$ with $\mathcal{Q} \in \mathrm{Sym}(d, \mathbb{R})$, by transforming $\mathcal{Q}_1$ into $\mathcal{Q}_2$ with the congruence matrix $\mathcal{Q}_1^{-1/2} \mathcal{Q}_2^{1/2} \in \mathrm{GL}(d, \mathbb{R})$.
Every ellipsoid can thus be transformed linearly into the unit ball.
- For the special linear group $\mathrm{SL}(d, \mathbb{R}) \subset \mathrm{GL}(d, \mathbb{R})$, the volume of the ellipsoid is the only invariant, and it is proportional to $\det(\mathcal{Q})^{-1/2}$.
- In contrast, for the orthogonal group $\mathrm{O}(d) \subset \mathrm{GL}(d, \mathbb{R})$, the lengths of the principal axes of the ellipsoid (there are d of them) are the invariants.

In comparison to the latter case, for $d = 2n$ and the symplectic group, there are just half as many invariants:

17.1 Lemma (Symplectic Normal Forms of Ellipsoids)
For every ellipsoid $\mathcal{E} \subset \mathbb{R}^{2n}$ in the canonical symplectic vector space $(\mathbb{R}^{2n}, \omega_0)$, there are unique real numbers $0 < r_1 \le \ldots \le r_n$ and a symplectic mapping $f \in \mathrm{Sp}(2n, \mathbb{R})$ such that $\mathcal{E} = f(\mathcal{E}_{r_1,\ldots,r_n})$ for the ellipsoid

$$\mathcal{E}_{r_1,\ldots,r_n} := \left\{(p, q) \in \mathbb{R}^{2n} \;\middle|\; \textstyle\sum_{k=1}^n \frac{p_k^2 + q_k^2}{r_k^2} \le 1\right\} .$$

[2] Its dimension is $n(2n + 3)$, because $\dim \mathrm{Sp}(E, \omega) = n(2n + 1)$ by Exercise 6.26.

Proof This follows from Lemma 6.29 and the normal form (6.3.3). □

Thus the linear symplectic images of the ellipsoids $B_r^{2n} = \mathcal{E}_{r,\ldots,r}$ are the best possible analogs of the Euclidean ball B_r^{2n} of radius $r > 0$, and for $n \geq 2$, not every ellipsoid can be so written with an appropriate radius r.

We consider the *symplectic cylinders* of the form

$$Z_R := \left\{(p,q) \in \mathbb{R}_p^n \times \mathbb{R}_q^n = \mathbb{R}^{2n} \mid p_1^2 + q_1^2 \leq R^2\right\}.$$

The question is now, when does there exist an $f \in \mathrm{ASp}(\mathbb{R}^{2n}, \omega_0)$ such that $f(B_r) \subset Z_R$?

17.2 Theorem (Gromov's Non-Squeezing Result in the Linear Case)
It is exactly for $r \leq R$ that a ball of radius r can be mapped affine symplectically into the cylinder Z_R.

Proof

- For $r \leq R$, the ball is mapped into the cylinder under $f = \mathrm{Id}$.

- The nonexistence part of the statement follows from its special case $r = 1$ by scaling. Let $f(x) = g(x) + a$ with $g \in \mathrm{Sp}(\mathbb{R}^{2n}, \omega_0)$ and $a = (a_1, \ldots, a_{2n})^\top \in \mathbb{R}^{2n}$. The left side of the condition

$$\max_{x \in S^{2n-1}} \left(f_1(x)^2 + f_{n+1}(x)^2\right) \leq R^2$$

is minimal for $a_1 = a_{n+1} = 0$.

The transpose $H = (h_1, \ldots, h_{2n})$ of the matrix representing g is also symplectic, and therefore $\|h_k\| \, \|h_{n+k}\| \geq \omega_0(h_k, h_{n+k}) = 1$. We conclude

$$\begin{aligned}\max_{x \in S^{2n-1}} \left(f_1(x)^2 + f_{n+1}(x)^2\right) &= \max_{x \in S^{2n-1}} \left(\langle h_1, x\rangle^2 + \langle h_{n+1}, x\rangle^2\right) \\ &\geq \max\{\|h_1\|^2, \|h_{n+1}\|^2\} \geq 1,\end{aligned}$$

by using the inequality $ab \leq \max\{a^2, b^2\}$ for $a := \|h_1\|$ and $b := \|h_{n+1}\|$. □

17.3 Remark (Squeezing for Volume Preserving Maps)
If one asks the analogous question for *affine volume preserving maps* (where we require in (17.1.1) that $g \in \mathrm{SL}(2n, \mathbb{R})$), then a corresponding imbedding is always possible for $n \geq 2$ degrees of freedom, because in this case, the cylinder Z_R (which itself can be viewed as the degenerate ellipsoid) has infinite volume. So the statement of the theorem refers to a specific property of symplectic mappings that goes beyond the conservation of volume. ◊

17.4 Corollary *The ellipsoid $\mathcal{E}_{r_1,\ldots,r_n}$ can be mapped affine symplectically into the cylinder Z_R if and only if $r_1 \leq R$.*

What is essential is that the *symplectic area* πr_1^2 of the ellipsoid is smaller than the symplectic area πR^2 of the cylinder. So the non-squeezing theorem measures a two-dimensional property of subsets of the symplectic phase space. Fundamental invariants of symplectic theory are 2-dimensional, whereas in Riemannian geometry, lengths of curves are 1-dimensional invariants.

We can also see this in the example of the *displacement energy* of a (compact) subset $K \subset M$ of a symplectic manifold (M, ω), introduced by Helmut Hofer:

$$e(K) \ := \ \inf_{\{H_t\}} \int_0^1 \big(\sup(H_t) - \inf(H_t)\big)\, \mathrm{d}t\,.$$

This quantity measures the energy needed to separate the set K from itself, namely $\phi_1(K) \cap K = \emptyset$, under a Hamiltonian flow $\{\phi_t\}_{t\in[0,1]}$ generated by $\{H_t\}_{t\in[0,1]}$.

17.5 Example (Displacement Energy)
We consider the simplest case of a flow generated by a linear, time independent Hamiltonian, and acting on an ellipsoid K. The ellipsoid is to be oriented parallel to the coordinate axes, i.e., $K = \Big\{(p,q) \in \mathbb{R}^{2n} \;\Big|\; \sum_{k=1}^n \frac{p_k^2}{r_k^2} + \frac{q_k^2}{s_k^2} \le 1\Big\}$.
The flow generated by the Hamiltonian

$$H_{\xi,\eta} : \mathbb{R}^{2n} \to \mathbb{R} \quad , \quad H_{\xi,\eta}(p,q) = \langle \xi, p\rangle - \langle \eta, q\rangle \qquad (\xi,\eta \in \mathbb{R}^n)$$

is the translation $\phi_t(p,q) = (p + t\eta, q + t\xi)$.
• Then separation ($\phi_t(K) \cap K = \emptyset$) is accomplished for all times

$$|t| > T(\xi,\eta) := 2\left(\sum_{k=1}^n \left(\frac{\eta_k^2}{r_k^2} + \frac{\xi_k^2}{s_k^2}\right)\right)^{-1/2}.$$

It is exactly when $(\eta,\xi)/2$ is a point on the surface of K that one has $T(\xi,\eta) = 1$.
• On the other hand, the energy employed is

$$\begin{aligned} \|H_{\xi,\eta}\| &:= \max\{H_{\xi,\eta}(x) \mid x \in K\} - \min\{H_{\xi,\eta}(x) \mid x \in K\} \\ &= \ 2\max\{H_{\xi,\eta}(x) \mid x \in K\}\,, \end{aligned}$$

where we have used the point symmetry of K. The maximization under the constraint $x \in K$ yields

$$\|H_{\xi,\eta}\| = 2\big(\langle \xi, r\rangle^2 + \langle \eta, s\rangle^2\big)^{1/2}.$$

Writing the components of the translation vector in polar coordinates,

$$\xi_k = \ell_k s_k \cos\varphi_k, \quad , \quad \eta_k = \ell_k r_k \sin\varphi_k \quad \text{with} \quad \ell_k := \sqrt{\frac{\eta_k^2}{r_k^2} + \frac{\xi_k^2}{s_k^2}}\,, \qquad (17.1.2)$$

one has $\|H_{\xi,\eta}\| = 2\big(\sum_{k=1}^n (\ell_k r_k s_k)^2\big)^{1/2}$.

• We now assume $T(\xi, \eta) = 1$, hence $\|\ell\|_2 = 2$. Letting $f_j = \min\{f_1, \dots, f_n\}$ where $f_k := r_k s_k$, the minimum of $\|H_{\xi,\eta}\|$ is $4 f_j$, taken on whenever $\ell_j = 2$ in (17.1.2), and $\ell_k = 0$ for all other k. Thus the displacement energy is proportional to the above-defined symplectic area of the ellipsoid. ◊

While the displacement energy assigns to sets like the ellipsoids a nonnegative number, this number does depend on the imbedding of the set into the symplectic manifold (which in the example of the ellipsoids was $(\mathbb{R}^{2n}, \omega_0)$); it is therefore not an intrinsic quantity.

As a consequence, in [EH], Ekeland and Hofer defined axiomatically, what is to be meant by an (intrinsic) symplectic capacity.

17.6 Definition
A mapping $c : (P, \omega) \mapsto [0, \infty]$ that assigns nonnegative numbers to symplectic manifolds is called a **symplectic capacity**, if it has the following properties:

Monotonicity: If there is a canonical transformation that imbeds (P_1, ω_1) into (P_2, ω_2), then $c(P_1, \omega_1) \leq c(P_2, \omega_2)$.
Scaling: $c(P, k\,\omega) = k\,c(P, \omega)$ for $k > 0$.
Nontriviality: $c(B_1^{2n}, \omega_0) > 0$ for the unit balls B_1^{2n} in $(\mathbb{R}^{2n}, \omega_0)$; $c(Z_1, \omega_0) < \infty$ for the symplectic cylinders Z_1 in $(\mathbb{R}^{2n}, \omega_0)$.

17.7 Theorem (Gromov's Nonsqueezing Result)
If there exists a symplectic capacity for which $c(B_1^{2n}, \omega_0) = c(Z_1, \omega_0)$, then the following result holds true:
A ball of radius r can be symplectically mapped into the cylinder Z_R of radius R if and only if $r \leq R$.

Proof
• If $r \leq R$, then B_r^{2n} is already a subset of Z_R.
• If there exists a symplectic imbedding $I : B_r^{2n} \to Z_R$, then

$$c(B_r^{2n}, \omega_0) \leq c(Z_R, \omega_0) = c(Z_1, R^{-2}\omega_0) = R^{-2} c(Z_1, \omega_0) = R^{-2} c(B_1^{2n}, \omega_0)\,,$$

which, in view of $c(B_r^{2n}, \omega_0) = r^{-2} c(B_1^{2n}, \omega_0)$, would contradict the positivity of $c(B_1^{2n}, \omega_0)$ if it were $r > R$. □

As a matter of fact, Gromov and Hofer-Zehnder have defined such capacities.

17.2 The Theorem by Poincaré-Birkhoff

Poincaré, who was the first to introduce qualitative methods into dynamics systematically, made a statement about area preserving mappings in his 'last geometric theorem'. This statement was proved by George David Birkhoff in 1925.

We consider, for the radii $0 < r_- < r_+ < \infty$ and the circle $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ a homeomorphism

$$H = (R, \Phi) : A \longrightarrow A \quad \text{of the annulus} \quad A := [r_-, r_+] \times S^1 \tag{17.2.1}$$

that maps the two boundary components $\{r_\pm\} \times S^1$ to themselves in an orientation-preserving manner. The mapping

$$\tilde{H} = (\tilde{R}, \tilde{\Phi}) : \tilde{A} \to \tilde{A}$$

that is semiconjugate to H with respect to the covering $\tilde{A} := [r_-, r_+] \times \mathbb{R} \to A$ has the periodicities $\tilde{R}(r, \varphi + 2\pi) = \tilde{R}(r, \varphi)$ and

$$\tilde{\Phi}(r, \varphi + 2\pi) = \tilde{\Phi}(r, \varphi) + 2\pi . \tag{17.2.2}$$

Assume the boundary components are turned in opposite directions, namely assume that there exist $\varphi_- < 0 < \varphi_+$ with

$$\tilde{\Phi}(r_\pm, \varphi) = \varphi + \varphi_\pm \qquad (\varphi \in \mathbb{R}). \tag{17.2.3}$$

The latter is mainly a requirement on H, but also on the choice of the lift $\tilde{H}$.

17.8 Theorem (Poincaré-Birkhoff)
If the homeomorphism H is area preserving[3] *then it has at least two fixed points in the interior of the annulus A.*

17.9 Remark The hypothesis that H be area preserving cannot be omitted. Counterexamples are the diffeomorphisms of the annulus given by

$$\tilde{H}_\varepsilon(r, \varphi) := \big(r - \varepsilon(r - r_+)(r - r_-)\,,\ \varphi + r - \tfrac{1}{2}(r_+ + r_-)\big) \qquad \big(|\varepsilon| < \tfrac{1}{r_+ - r_-}\big).$$

They are without fixed points when $\varepsilon \neq 0$, but are area preserving only for $\varepsilon = 0$. ♢

Proof We prove the theorem under the simplifying assumption that H is twice continuously differentiable. In the paper [Bi2] by George David Birkhoff, one can find a proof that does not make this simplifying hypothesis. The mapping

$$\Psi : A \to \mathbb{R} \quad , \quad \Psi(r, [\varphi]) := \tilde{\Phi}(r, \varphi) - \varphi$$

is well-defined by (17.2.2), and it measures the angle difference between the image of a point under H and the point itself. In particular, one has $\Psi(r_\pm, [\varphi]) = \varphi_\pm$ from (17.2.3), and therefore $0 \in \Psi(A)$.

[3] We take the Lebesgue measure on A, respectively the area form $dr \wedge d\varphi$, but one may also take the measure $r\, dr \wedge d\varphi$ that arises from polar coordinates on $\mathbb{R}^2$.

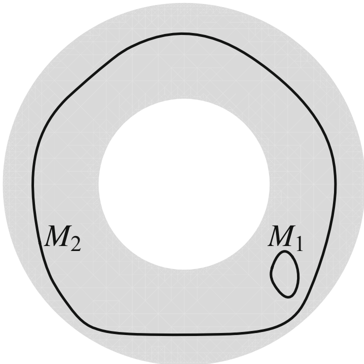
- If 0 is a regular value of Ψ, then the pre-image $M := \Psi^{-1}(0) \subset A$ is a 1-dimensional compact submanifold. It therefore consists of finitely many connected components M_i, which in turn are diffeomorphic to the circle.

  The restrictions $\Pi_i : M_i \to S^1$ of the projection $\Pi : A \to S^1$, $(r, \varphi) \mapsto \varphi$ have therefore mapping degrees $\deg(\Pi_i)$ (see page 131), which we may assume to be nonnegative by choosing the orientation of M_i.

  A mapping degree larger than 1 cannot occur, because otherwise M_i would have to intersect itself. The M_i are Jordan curves. Thus $A \setminus M_i$ consists of exactly two connected components because of the Jordan curve theorem.[4] But there must be an index j with $\deg(\Pi_j) = 1$, because otherwise $\{r_-\} \times S^1$ and $\{r_+\} \times S^1$ would lie in the same component.

  Intersections of M_j with $H(M_j)$ are fixed points of A, because under the mapping by A, only the first component of a point $(r, \varphi) \in M$ can change.

  Such intersections do exist, because otherwise one connected component U of $A \setminus M_j$ would be mapped under H into a proper subset $H(U)$ of itself. This would be in contradiction to the area preserving property, because then the area of this subset would be smaller as well.



  However, it there exists one point of intersection of M_j with $H(M_j)$, then there also exists a second such point.

- If however 0 is not a regular value of Ψ, then we consider the perturbations

$$H_\varepsilon : A \to A \quad , \quad H_\varepsilon := R_\varepsilon \circ H \quad \text{with} \quad R_\varepsilon(r, \varphi) := (r, \varphi - \varepsilon)$$

  of H. For all $\varepsilon \in (\varphi_-, \varphi_+)$, the mapping H_ε satisfies the hypotheses of the theorem. By Sard's theorem (see page 319), there exists a sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ converging to 0, for which H_{ε_n} also satisfies the regularity hypothesis from the first part of the proof. Thus there exists a sequence of fixed points x_n of H_{ε_n}. By compactness of the annulus, this sequence has a convergent subsequence. Its limit point $x \in A$ is a fixed point of H.

  A refined argument would show that there are indeed two fixed points. □

[4] **Jordan Curve Theorem:** The complement of the image $c(S^1) \subset \mathbb{R}^2$ of a simple closed curve $c : S^1 \to \mathbb{R}^2$ in the plane consists of exactly two connected components, one of which is bounded, and the other of which is unbounded. $c(S^1)$ is the common boundary of both connected components.

A particularly simple class of mappings (17.2.1) consists of the *monotone twist maps*. These mappings are continuously differentiable, with partial derivative $D_1\Phi > 0$. Therefore, by the implicit function theorem, the zero set $M = \Psi^{-1}(0) \subset A$ is the graph of a function of the angle.

17.10 Example (Monotone Twist Maps)

1. **(Standard Map)** The area preserving standard maps considered in Sect. 15.6,

$$F_\varepsilon(x, y) = \big(x + y + \varepsilon \sin(2\pi x),\ \ y + \varepsilon \sin(2\pi x)\big)$$

satisfy the monotone twist property $\mathrm{D}_1\Phi > 0$. Here x is considered as the angle variable φ and $y \in \mathbb{R}$ as the radius r, such that

$$\Phi(r, \varphi) = \varphi + r + \varepsilon \sin(2\pi\varphi)\,.$$

In these coordinates, admittedly, the condition of invariance and opposite direction of rotation (17.2.3) of the boundary circles is not satisfied when $\varepsilon \neq 0$. However, one can consider the domain between two invariant KAM-tori, and this idea is at the basis of the real potential for applying the Poincaré-Birkhoff theorem; then one can conclude the existence of two fixed points in this domain.

On the other hand, in this example, one can also see directly that the points $(x, y) = (0, 0)$ and $(x, y) = (1/2, 0)$ are fixed points of the standard maps F_ε.

One can also calculate that the linearization of F_ε at these fixed points is of the form $\left(\begin{smallmatrix} 1+2\pi\varepsilon & 1 \\ 2\pi\varepsilon & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1-2\pi\varepsilon & 1 \\ -2\pi\varepsilon & 1 \end{smallmatrix}\right)$ respectively, such that for $\varepsilon \neq 0$, one is hyperbolic, the other elliptic, in the sense of the classification of matrices in $\mathrm{SL}(2, \mathbb{R})$ (Exercise 6.26).

The Figure 15.6.1 give rise to the conjecture (and this can be proved) that a chaotic zone is formed around the hyperbolic fixed point, whereas the elliptic fixed point is surrounded by invariant tori.

All these are typical properties of twist maps.

2. **(Convex Billiard)**
A simple closed curve $c : S^1 \to \mathbb{R}^2$ defines a bounded domain in $\mathbb{R}^2$ according to the JordanAn achromatic lens or, more precisely, achromatic curve theorem that we have just used. This domain can be interpreted as a billiard table. The situation is simplest if the curve is smooth, regular, and with positive curvature.[5]
We normalize the length of the curve to 2π.

[5]If c is parametrized by arclength (which is possible for regular curves and will be assumed in the sequel), then positive curvature means $\big\langle c''(t), \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) c'(t)\big\rangle > 0$.

Then the trajectory of a billiard consists of a polygonal path whose corners are on the boundary $\mathcal{C} := c(S^1)$ of the billiard table and for which the directions with respect to the normal vector $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) c'(t)$ to the curve change according to the rule 'incoming angle = − outgoing angle'. The segment is parametrized by the arclength t of the point of collision and the outgoing angle (or more practically its sine, u). The phase space of the discretized dynamics is thus an annulus of the form $A := [-1, 1] \times S^1$.

Due to the positive curvature of the boundary, each successive collision with the boundary is transversal, and therefore, the mapping $\Phi : A \to A$ is smooth by the implicit function theorem.

Transversality only fails in the case of a direction that is tangential to $\mathcal{C}$, namely when $u = \pm 1$. In this case, (u, t) is a fixed point of Φ, and therefore one of the conditions of Theorem 17.8 is satisfied.

For the same reason however, condition (17.2.3) about the opposite rotation of the boundary components is violated. The statement of the theorem is also not valid for this example, because for an angle different from $\pm\pi$, the next point of collision is different from the given one, so there is no fixed point in $\mathring{A} = (-1, 1) \times S^1$.

Nevertheless, the Poincaré-Birkhoff theorem can be applied to the billiard problem. As the angle difference $\varphi_+ - \varphi_-$ (see (17.2.3)) equals 2π, the n^{th} iterate Φ^n corresponds to an angle difference of $2\pi n$. Therefore, for all $n \geq 2$, the mapping Φ^n has fixed points in $\mathring{A}$. They belong to n-periodic orbits of Φ.

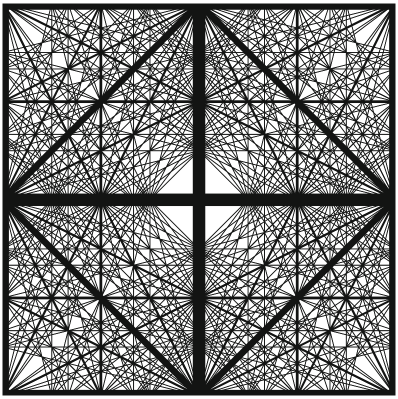
Since Φ^m and Φ^n cannot have a common fixed point in $\mathring{A}$ when m and n are relatively prime, Φ has infinitely many periodic orbits. ◊

17.3 The Arnol'd Conjecture

George David Birkhoff was already looking for a generalization of his theorem for higher phase space dimensions. Such a generalization was successful in the past years with the proof of the Arnol'd conjecture.

17.11 Remark (Arnol'd Diffusion) At first, it is not even clear *how* such a statement could even look. Let us consider as domain of applicability for the Poincaré-Birkhoff theorem the phase space portrait of a Hamiltonian system with two degrees of freedom, as in Figure 15.4.3. As the energy surface is 3-dimensional, the 2-dimensional KAM tori have codimension 1. Thus two such KAM tori can bound a 3-dimensional domain. After discretizing the dynamics by means of a Poincaré surface, we obtain an annulus between two invariant circles.

Now, since for n degrees of freedom, the n-dimensional KAM tori have codimension $n-1$ in the energy shell, finitely many of them will not bound a domain, when $n > 2$. Instead, it is possible in this case that initial values that do not lie on an invariant torus belong to trajectories whose action variable changes with time in a manner similar to a random walk.



Correspondingly, one has a complement of the Diophantine set in the spheres $\|\omega\| =$ const that is connected. For $n = 3$ and $\omega_3 = 1$, such a set is depicted on the right (in black). Compare this with Figure 15.3.1 for $n = 2$ on page 408.

This phenomenon, called *Arnol'd diffusion* could be proved for some Hamiltonian systems; for instance in the case of motion in a periodic potential, worked out by KALOSHIN and LEVI [KL]. See also Chapter 6 of LICHTENBERG and LIEBERMAN [LL]. $\diamondsuit$

So we cannot expect a generalization in this direction. However, the following, alternative proof of the Poincaré-Birkhoff theorem for the Billard problem yields a more promising perspective.

17.12 Example (Periodic Orbits of Convex Billards)
In Example 17.10.2, we parametrized the collision data by the phase space coordinates $(u, t) \in A = [-1, 1] \times S^1$; accordingly, the discrete dynamics $H = (R, \Phi) : A \to A$ mapped a point (u_0, t_0) into (u_1, t_1).

The monotone twist condition $D_1\Phi > 0$ allows to calculate the initial direction u_0 from the pair (t_0, t_1) of points on the boundary $\mathcal{C}$. Namely, if we use

$$S : S^1 \times S^1 \to \mathbb{R} \quad , \quad S(t_0, t_1) = -\|c(t_1) - c(t_0)\|$$

as the generating function for a canonical transformation, then S is a smooth function except on the diagonal $\Delta \subset S^1 \times S^1$ of the torus; and the diagonal points $(t, t) \in \Delta$ correspond to improper billiard trajectories that are tangential to the boundary. The partial derivatives are

$$D_1S(t_0, t_1) = \langle c'(t_0), c(t_1) - c(t_0)\rangle / \|c(t_1) - c(t_0)\| = \sin\varphi_0 = u_0$$

and analogously $D_2S(t_0, t_1) = u_1$. Thus the critical points (t_0, t_1) of S correspond to $u_0 = u_1 = 0$, which is a segment of the billiard trajectory that hits the boundary $\mathcal{C}$ orthogonally at both ends. Iteration leads to a 2-periodic orbit in this case.

We conclude the existence of such an orbit from the fact that S has a negative average, whereas S vanishes on Δ. This minimum is taken on twice because of the symmetry $S(t_1, t_0) = S(t_0, t_1)$.

Another periodic orbit corresponds to a (possibly degenerate) saddle point of S. One can conclude its existence from the fact that S vanishes on the diagonal. Thus, if we cut the torus open along the diagonal, S corresponds to a function on the annulus A that takes the value zero on both boundary circles $\partial_\pm A$ and that is negative in the interior. The saddle point can be found, for example, by considering the minimax problem

$$\max_{d\in\mathcal{D}} \min_{x\in[0,1]} S\big(d(x)\big) < 0\,,$$

with $\quad \mathcal{D} := \big\{d : [0,1] \to A \mid d \text{ continuous},\ d(0) \in \partial_- A,\ d(1) \in \partial_+ A\big\}\,.$ ◊

17.13 Exercise (Elliptic Billiard)
For an elliptic billiard table with semiaxes $0 < a_1 < a_2$ and boundary

$$\mathcal{C} := \{x \in \mathbb{R}^2 \mid (x_1/a_1)^2 + (x_2/a_2)^2 = 1\}\,,$$

show that the discrete dynamics has exactly two geometrically distinct periodic trajectories of minimal period 2. ◊

The minimal number of critical points is therefore two in this case. In connection with such examples, ARNOL'D made the following conjecture in the 1960s, see Appendix 9 in [Ar2].

Arnol'd Conjecture: A Hamiltonian symplectomorphism $\Phi : P \to P$ (see Definition 10.27) of a compact symplectic manifold P

- has at least as many fixed points as a smooth function $S : P \to \mathbb{R}$ has critical points.
- If its fixed points are nondegenerate, their number is at least as large as the minimal number of critical points of a Morse function $S : P \to \mathbb{R}$.

We have used here the following notions (see also Appendix G):

17.14 Definition *For a differentiable manifold P,*

- *a fixed point $x \in P$ of a mapping $\Phi \in C^1(P, P)$ is called* **nondegenerate** if the linear mapping $\mathrm{D}\Phi(x) - \mathrm{Id}_{T_x P} : T_x P \to T_x P$ is bijective;
- a function $f \in C^2(P, \mathbb{R})$ is called a **Morse function** if all critical points x of f are nondegenerate, i.e., if the Hessian $\mathrm{D}^2 f(x)$ is nondegenerate as a bilinear form in the sense of Definition 6.12.

17.15 Remark (Arnol'd Conjecture)

1. Since the minimal numbers of critical points addressed in the conjecture can be calculated practically (see Appendix G), the inequalities thus obtained yield an approach to periodic orbits of Φ.
2. If the Hamiltonian symplectomorphism $\Phi : P \to P$ is the time-1 flow of the Hamiltonian differential equation for a Hamiltonian $H : P \to \mathbb{R}$, then the Arnol'd conjecture for Φ is obviously true because the equilibria of the flow correspond to critical points of H. This also shows that the claimed inequalities are optimal if P carries a perfect Morse function (see page 580).

3. In general, in order to get the existence of fixed points, it is necessary to assume that the symplectomorphism Φ is Hamiltonian.

   This is because the 2-torus $P := \mathbb{T}^2$ with area form ω is a compact symplectic manifold, and every translation $\Phi_a : P \to P,\ x \mapsto x + a$ is symplectic. But it is free of fixed points except in the Hamiltonian case $\Phi_0 = \mathrm{Id}_P$ (see Example 10.29.3).
4. A Morse function $f \in C^2(P, \mathbb{R})$ is called a *Morse-Smale function* if (other than in the figure on page 581) any intersection of the stable manifold of a critical point p with the unstable manifold of a critical point q is transversal. Then Morse theory can be built on relative indices of (p, q) defined means of the transversal intersections; see M. SCHWARZ [Schw]. This allows the *Floer theory*, named after ANDREAS FLOER, for infinite dimensional P, see [Bo2] by R. BOTT, [McD] by D. MCDUFF and Chapter 6 in [Jo] by J. JOST. ◊

17.16 Literature Useful for a thorough study are the article [Ar5] by ARNOL'D, which defines the area, the textbooks by MCDUFF and SALAMON [MS], by HOFER and ZEHNDER [HZ, Zeh], and the literature quoted therein. ◊

Appendix A
Topological Spaces and Manifolds

A.1 Topology and Metric

A subset of $\mathbb{R}$ is called *open* if it is the union of open intervals $(a, b) \subseteq \mathbb{R}$. This definition makes $\mathbb{R}$ into a topological space:

A.1 Definition

1. *A **topological space** is a pair $(M, \mathcal{O})$, consisting of a set M and a set $\mathcal{O}$ of subsets of M (these subsets will be called **open sets**) such that the following properties are satisfied:*
   1. *$\emptyset$ and M are open,*
   2. *arbitrary unions of open sets are open,*
   3. *the intersection of any two open sets is open.*
2. *$\mathcal{O}$ is called a **topology** on M.*
3. *If $\mathcal{O}_1$ and $\mathcal{O}_2$ are topologies on M and $\mathcal{O}_1 \subseteq \mathcal{O}_2$, then we call $\mathcal{O}_1$ **coarser** or **weaker** than $\mathcal{O}_2$, and $\mathcal{O}_2$ **finer** or **stronger** than $\mathcal{O}_1$.*

A.2 Example 1. The *discrete topology* 2^M (the power set) is the finest topology on a set M, and the *trivial topology* $\{M, \emptyset\}$ is the coarsest topology on M. Topological spaces $(M, 2^M)$ are called *discrete*.

2. If $N \subseteq M$ is a subset of the topological space $(M, \mathcal{O})$, then

$$\{U \cap N \mid U \in \mathcal{O}\} \subseteq 2^N \tag{A.1.1}$$

   defines a topology on N, called the *subspace topology*, *relative topology*, *induced topology*, or *trace topology*. For example, for the subset $N := [0, \infty)$ of $\mathbb{R}$, all subintervals of the form $[0, b) \subset N$ with $b \in N$ are open in N, but not open in $\mathbb{R}$.
3. If $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ are disjoint topological spaces (i.e., $M \cap N = \emptyset$), then the union $M \cup N$ with $\{U \cup V \mid U \in \mathcal{O}_M,\ V \in \mathcal{O}_N\} \subseteq 2^{M \cup N}$ is a topological space, called the *sum space*.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

4. Let $(M, \mathcal{O}_M)$ be a topological space and $f : M \to N$ surjective. The *quotient topology* induced by f on N is the topology

$$\left\{V \subseteq N \mid f^{-1}(V) \in \mathcal{O}_M\right\} \subseteq 2^N. \qquad \Diamond$$

A.3 Theorem *Let $\mathcal{F}$ be an arbitrary family of subsets of a set M.*

- *Then there is a unique coarsest topology $\mathcal{O}(\mathcal{F})$ on M satisfying $\mathcal{F} \subseteq \mathcal{O}(\mathcal{F})$.*
- *$\mathcal{O}(\mathcal{F})$ is then called the topology **generated by** $\mathcal{F}$.*

In many cases, the topology is *generated by a metric*.

A.4 Definition

- *A **metric space** is a pair (M, d), consisting of a set M and a function $d : M \times M \to [0, \infty)$, such that the following properties hold for all $x, y, z \in M$:*

(a) $d(x, y) = 0 \Longleftrightarrow x = y$ *(Positivity)*
(b) $d(x, y) = d(y, x)$ *(Symmetry)*
(c) $d(x, z) \le d(x, y) + d(y, z)$ *(Triangle Inequality).*

- *d is called a **metric**, $d(x, y)$ the **distance between** x **and** y.*
- *For $x \in M$ and $\varepsilon > 0$, the set*

$$U_\varepsilon(x) := \{y \in M \mid d(y, x) < \varepsilon\}$$

*is called the (open) ε-**neighborhood of** x in M.*

A.5 Example 1. The set $B := \{0, 1\}$ designates one *bit*. We consider sequences of $n \in \mathbb{N}$ bits, i.e., elements[1] of B^n. Their *Hamming distance* is given by the metric $d : B^n \times B^n \to \{0, 1, \dots, n\}$,

$$d\big((b_1, \dots, b_n), (c_1, \dots, c_n)\big) := \Big|\big\{i \in \{1, \dots, n\} \mid b_i \neq c_i\big\}\Big|,$$

namely by the number of positions in which the bit sequences differ. This metric is used in information theory.

2. If (M, d) is a metric space and N a subset of M, then (N, d_N) is again a metric space with the metric ‘restricted’[2] to N,

$$d_N : N \times N \to \mathbb{R} \quad , \quad d_N(x, y) := d(x, y).$$

For instance, for $M := \mathbb{C}$ with the metric $d(x, y) := |x - y|$, the set $S^1 \subset \mathbb{C}$ is geometrically speaking the *circle* of radius 1 about the origin. The distance $d_{S^1}(x, y)$ between two points on the circle is then the length of the straight line segment between the points x and y. A different, sensible metric on S^1 is given as the (minimal) difference in angles of x and y.

[1] We will write them as row vectors.

[2] More precisely, it is the mapping $d : M \times M \to \mathbb{R}$ that is restricted to $N \times N$.

3. On the vector space $\mathbb{R}^n$, the length of a vector v is defined by the *Euclidean norm*

$$\|v\| := \sqrt{\sum_{i=1}^n v_i^2} \qquad \left(v = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n\right).$$

With $d(x, y) := \|y - x\|$, one obtains from it a metric on $\mathbb{R}^n$, called the *Euclidean metric*.

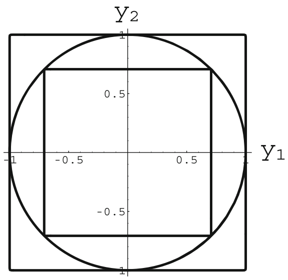
4. On the vector space $\mathbb{R}^n$, one can also define a metric in terms of the *maximum norm*

$$\|v\|_\infty := \max(|v_1|, \ldots, |v_n|) \qquad \left(v = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n\right).$$

$d_\infty(x, y) := \|x - y\|_\infty$, and d_∞ satisfies

$$d_\infty(x, y) \le d(x, y) \le \sqrt{n}\, d_\infty(x, y) \quad (x, y \in \mathbb{R}^n);$$

see the figure for $n = 2$ and $x = 0$.
The 'unit ball' $\{v \in \mathbb{R}^n \mid \|v\|_\infty \le 1\}$ with respect to the maximum norm is an n-dimensional cube of side length 2 that is parallel to the axes and centered in the origin. ♢



A.6 Definition *For a metric space (M, d), the set*

$$\mathcal{O}(d) := \left\{V \subseteq M \mid \forall x \in V\ \exists \varepsilon > 0 : U_\varepsilon(x) \subseteq V\right\}. \tag{A.1.2}$$

is called the (metric) ***topology on*** (M, d).

A.7 Theorem
$\left(M, \mathcal{O}(d)\right)$ *is a topological space, and the ε-neighborhoods $U_\varepsilon(x)$ are open.*

A.8 Remark (Topological Vector Spaces) Frequently, different metrics generate the same topology. This is in particular true for such metrics on vector spaces as arise from equivalent norms[3] (as in Examples A.5.3 and A.5.4). For in this case, one can find for every ε-ball around x with respect to one norm an ε/c-ball around x with respect to the other norm, that is contained in the former ball.

As all norms on a finite dimensional vector space are equivalent, we can talk about *the* topology of $\mathbb{R}^n$ or $\mathbb{C}^n$ (however, there are other topologies on these spaces, but they do not arise from norms, see Example A.2.1). On an infinite dimensional $\mathbb{K}$-vector space V, there are many different topologies $\mathcal{O}$ that are generated by norms.

In any case, addition and multiplication by scalars are continuous operations, so $(V, \mathcal{O})$ is called a *topological vector space*. ♢

---

[3]**Definition:** Two norms $\|\cdot\|_I, \|\cdot\|_{II} : V \to \mathbb{R}$ are called **equivalent** if there exists a number $c \ge 1$ with

$$c^{-1}\|v\|_I \le \|v\|_{II} \le c\|v\|_I \qquad (v \in V).$$

We expand our topological vocabulary by generalizing notions that are familiar from the space $\mathbb{R}$:

A.9 Definition *Let $(M, \mathcal{O})$ be a topological space.*

- *$A \subseteq M$ is called* ***closed*** *if A is the complement of an open set: $M \setminus A \in \mathcal{O}$.*
- *$U \subseteq M$ is called a* ***neighborhood*** *of $x \in M$ if there exists an open set V with $x \in V \subseteq U$.*
- *For $A \subseteq M$ and $x \in M$, we call x an* ***interior*** *resp.* ***exterior*** *resp.* ***boundary point of*** *A, depending on whether A or $M \setminus A$ or neither of the two sets is a neighborhood of x.*
  - *$\mathring{A} := \{x \in M \mid x \text{ is interior point of } A\}$ is called the* ***interior of*** *A.*
  - *$\overline{A} := \{x \in M \mid x \text{ not exterior point of } A\}$ is called the* ***closure*** *or the* ***closed hull of*** *A.*
  - *$\partial A := \{x \in M \mid x \text{ boundary point of } A\}$ is called the* ***boundary of*** *A.*
- *$x \in M$ is called* ***cluster point*** *of the subset $A \subseteq M$ if the set $U \cap (A \setminus \{x\})$ is not empty for any neighborhood U of x.*
- *$N \subseteq M$ is called* ***dense*** *if $\overline{N} = M$.*
- *$N \subseteq M$ is called* ***nowhere dense*** *if the interior of $\overline{N}$ is empty.*

A.10 Example

1. For $A := (0, 1] \subseteq \mathbb{R}$, we have $\mathring{A} = (0, 1)$, $\overline{A} = [0, 1]$, and $\partial A = \{0, 1\}$.
2. $\mathbb{Q}$ is dense in $\mathbb{R}$, and $\mathbb{Z}$ is nowhere dense in $\mathbb{R}$. ◊

A.11 Definition *Let $(M, \mathcal{O})$ be a topological space.*

- *A family $(U_i)_{i \in I}$ of $U_i \in \mathcal{O}$ is called an* ***open cover*** *of M if*

$$\bigcup_{i \in I} U_i = M.$$

- *$(M, \mathcal{O})$ is called* ***compact*** *if every open cover $(U_i)_{i \in I}$ has a* ***finite subcover****, i.e., an open cover $(U_j)_{j \in J}$ with a finite index set $J \subseteq I$.*
- *$(M, \mathcal{O})$ is called* ***locally compact*** *if all $m \in M$ have a compact neighborhood.*
- *$(M, \mathcal{O})$ is called a* ***Hausdorff space*** *if for all $x \neq y \in M$, there exist disjoint neighborhoods U_x of x and U_y of y.*
- *$(M, \mathcal{O})$ is called* ***paracompact*** *if it is Hausdorff and for every open cover $\{U_i\}_{i \in I}$, there exists an open cover $\{V_j\}_{j \in J}$ that is*

(a) a ***refinement of*** *$\{U_i\}_{i \in I}$, i.e., $\forall j \in J\ \exists i \in I : V_j \subseteq U_i$,*
(b) and that is ***locally finite****, i.e., for all $x \in M$ there exists a neighborhood U of x for which the index set $\{j \in J \mid V_j \cap U \neq \emptyset\}$ is finite.*

A.12 Example (Topological Notions)

1. Finite dimensional real and complex vector spaces V are locally compact and paracompact. Subsets of these V are compact if and only if they are closed and bounded. (The latter is called the Heine-Borel theorem.)
   However, $\mathbb{Q}$ with the relative topology from $\mathbb{R}$ is not locally compact.

2. As in a metric space (M, d), all points $x \neq y \in M$ have a positive distance, the topological space $\big(M, \mathcal{O}(d)\big)$ is Hausdorff.
   A topological space $(M, \mathcal{O})$ is called *metrizable* if there exists a metric d on M for which $\mathcal{O} = \mathcal{O}(d)$. The example of spaces that are not Hausdorff shows that not every topological space is metrizable.
3. Infinite-dimensional Banach spaces are not locally compact, but are paracompact (as is every metric space).
4. An example of a Hausdorff space that is not paracompact is the so-called 'long line', see HIRSCH [Hirs], Ch. 1.1, Exercise 9. Usually, spaces that fail to be paracompact do not occur in problems of mechanics. ♢

A.13 Definition • *A* ***partition of unity*** *on a topological Hausdorff space* $(M, \mathcal{O})$ *is a family* $(f_i)_{i\in I}$ *of continuous (see Definition A.17) functions* $f_i : M \to [0, 1]$ *such that each* $x \in M$ *has a neighborhood* U *such that the index set* $\{j \in I \mid f_j\restriction_U \neq 0\}$ *is finite and* $\sum_{i\in I} f_i(x) = 1$.

- *A partition of unity* $(f_i)_{i\in I}$ *is called* ***subordinate to*** *an open cover* $(U_i)_{i\in I}$ *of* M *if* $supp(f_i) \subseteq U_i$ *for all* $i \in I$.

Partitions of unity are useful because they permit to combine objects that are defined *locally*, i.e., on appropriate neighborhoods U_i, into a *globally* defined object, i.e., one that is defined on all of M, and to do so in a convex manner.
The following theorem therefore clarifies the significance of the notion 'paracompact':

A.14 Theorem *A topological Hausdorff space is paracompact if and only if it has, for every open cover, a subordinate partition of unity.*

A.15 Definition

- *A sequence* $(x_n)_{n\in\mathbb{N}}$ *in a metric space* (M, d) *is called a* ***Cauchy sequence*** *if for every* $\varepsilon > 0$, *there exists a bound* $N_0 \equiv N_0(\varepsilon) \in \mathbb{N}$ *such that*

$$d(x_m, x_n) < \varepsilon \qquad \big(m, n \geq N_0(\varepsilon)\big).$$

- (M, d) *is called* ***complete*** *if every Cauchy sequence* ***converges*** *to some* $x \in M$, *i.e., for all* $\varepsilon > 0$, *there exist a bound* $N(\varepsilon) \in \mathbb{N}$ *such that*

$$x_n \in U_\varepsilon(x) \qquad \big(n \geq N(\varepsilon)\big).$$

A.16 Theorem *Every compact metric space is complete.*

A.17 Definition
A mapping $f : M \to N$ *of the topological spaces* $(M, \mathcal{O}_M)$ *and* $(N, \mathcal{O}_N)$

- *is called* ***continuous*** *if the pre-images of open sets are open (*$f^{-1}(V) \in \mathcal{O}_M$ *provided* $V \in \mathcal{O}_N$*);*
- *is called a* ***homeomorphism*** *if* f *is bijective and continuous and* $f^{-1} : N \to M$ *is continuous.*

- *The topological spaces* $(M, \mathcal{O}_M)$ *and* $(N, \mathcal{O}_N)$ *are called* ***homeomorphic*** *if there exists a homeomorphism* $f : M \to N$.

Homeomorphic topological spaces cannot be distinguished within the vantage point of topology. An example is $\mathbb{R}$ and the interval $(-1, 1)$, with the homeomorphism $\tanh : \mathbb{R} \to (-1, 1)$.

For the cartesian product $M := \prod_{i\in I} M_i$ of sets M_i, we define a topology, based on topologies $\mathcal{O}_i$ of the factors M_i, by means of the *canonical projections*

$$p_j : M \to M_j \quad , \quad (m_i)_{i\in I} \mapsto m_j \qquad (j \in I)\,,$$

as follows:

A.18 Definition

- *The* ***product topology*** $\mathcal{O}$ *on* M *is the coarsest topology with respect to which all canonical projections* $p_i \quad (i \in I)$ *are continuous.*
- *The topological space* $(M, \mathcal{O})$ *is then called the* **product space** *of the* $(M_i, \mathcal{O}_i)$.

A.19 Theorem (Tychonoff)[4] *Arbitrary products of compact topological spaces (with the product topology) are compact.*

Proof: See for instance JÄNICH [Jae], Chapter X. □

$U \subseteq M$ is open if and only if U is the union of (possibly infinitely many) finite intersections of sets of the form $p_i^{-1}(O_i)$ with $O_i \in \mathcal{O}_i$.

In general, not all cartesian products of open sets are open. They are however open if the index set I is finite.

A.20 Definition *Let* $(M, \mathcal{O})$ *be a topological space.*

- $(M, \mathcal{O})$ *is called if* ***connected*** *if the only sets that are both open and closed in* M *are the empty set and* M *itself.*
  Otherwise, $(M, \mathcal{O})$ *is called* ***disconnected***.
- $N \subseteq M$ *is called* ***connected*** *if* N *is connected in the relative topology (A.1.1) of* $\mathcal{O}$.
- *The maximal connected subsets* $N \subseteq M$ *are called the* ***connected components*** *of* M.
- $(M, \mathcal{O})$ *is called* ***totally disconnected*** *if all connected components consist (each) of a single point.*
- $N \subseteq M$ *is called* ***discrete*** *if the relative topology is discrete (see Exercise A.2.1).*

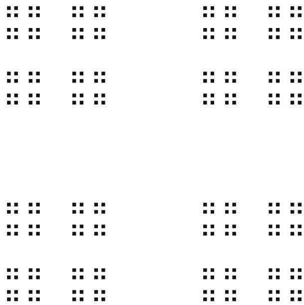
---

[4] Added in translation: this English transliteration of the Russian mathematician's name seems to be customary for historical reasons, even though Tikhonov would be more in line with transliteration standards and phonetic similarity.

A.21 Remark

1. Connected components $N \subseteq M$ are closed in $(M, \mathcal{O})$, but need not be open. For instance the rational numbers (as single-element sets) are the connected components of $\mathbb{Q}$.
2. The connected components N_i $(i \in I)$ of $(M, \mathcal{O})$ form a partition of M, i.e., $\cup_{i \in I} N_i = M$ and $N_i \cap N_j = \emptyset$ for $i \neq j \in I$.
3. A metrizable (see Example A.12.2) space $(M, \mathcal{O})$ is called a (topological) **Cantor set** if the space is non-empty, compact, and totally disconnected, and each point in M is a cluster point of M.

   All Cantor sets are homeomorphic to each other and also to the Cantor 1/3 set (see Exercise 2.5). This also applies, for example, to the cartesian product of the Cantor 1/3 set with itself, as depicted on the right. ◊



A.22 Definition

- *Two continuous mappings $f_0, f_1 : M \to N$ are called **homotopic** if there exists a continuous mapping*

$$h : M \times [0,1] \to N \quad \textit{with} \quad h(m,i) = f_i(m) \quad (m \in M,\ i \in \{0,1\}).$$

  *In this case, h is called a **homotopy** from f_0 to f_1.*
- *Two topological spaces M, N are called **homotopy equivalent** if there exist continuous mappings $f : M \to N$ and $g : N \to M$ such that $g \circ f$ is homotopic to Id_M and $f \circ g$ is homotopic to Id_N.*
- *A topological space M is called **contractible** if M is homotopy equivalent to a singleton (i.e., a set consisting of one point).*
- *A continuous mapping $c : I \to M$ of an interval into a topological space is called a **curve** or **path**. $t \in I$ is then called the* **parameter** *of c, the image $c(I) \subseteq M$ also the* **trace of** *c.*
- *Two curves $c_0, c_1 : I \to M$ in a topological space $(M, \mathcal{O})$, where $I := [a,b]$, that have the same initial and end points ($c_0(a) = c_1(a)$ and $c_0(b) = c_1(b)$) are called **homotopic** relative to these points if there exists a homotopy $h : I \times [0,1] \to M$ from c_0 to c_1 for which*

$$h\restriction_{\{a\}\times[0,1]} = c_0(a) \quad \textit{and} \quad h\restriction_{\{b\}\times[0,1]} = c_0(b).$$

- *M is called **simply connected** if any two curves in M with the same initial and end points are homotopic relative to these points.*

- *For an* $m \in M$*, the* ***fundamental group*** $\pi_1(M, m)$ ***of*** M *(at* m*) is the set of homotopy classes of curves that begin and end at* m*.*

A.23 Remark

1. Intuitively, the homotopy h deforms the curve c_0 continuously into c_1 while the initial and end points are held fixed. Namely, if one defines, for $s \in [0, 1]$, the curve
$$c_s : I \to M \quad , \quad c_s(t) := h(t, s),$$
then the curves c_s coincide for $s = 0$ or $s = 1$ with the previously defined curves, and $c_s(a) = c_0(a) = c_1(a)$, $c_s(b) = c_0(b) = c_1(b)$ for $s \in [0, 1]$.
2. As we have already used tacitly when defining the fundamental group, homotopy equivalence is an equivalence relation.
3. The fundamental group does have the structure of a group, with composition of curves defining the group operation.
If the base points m_1 and m_2 are in the same connected component of M, the groups $\pi_1(M, m_i)$ are isomorphic. Such an isomorphism is induced by a curve with initial point m_1 and end point m_2. Therefore, if M is connected, one may simply talk about *the* (abstract) fundamental group $\pi_1(M)$ of M. ♢

A.24 Example (Connectedness)

1. The punctured space $M := \mathbb{R}^{n+1} \setminus \{0\}$ is homotopy equivalent to the sphere $N := S^n$. This can be seen by taking for $f : M \to N$ the radial projection $x \mapsto x/\|x\|$, and for $g : N \to M$ the imbedding. This then implies that $\pi_1(M) = \pi_1(S^n)$.
All spheres S^n except $S^0 = \{-1, 1\}$ are connected.
S^n is simply connected if and only if $n \geq 2$. In that case, the fundamental group of S^n is trivial. Moreover, $\pi_1(S^1) \cong \mathbb{Z}$.
2. Convex subsets $U \subseteq \mathbb{R}^n$ are simply connected, because for any two curves $c_0, c_1 : [a, b] \to U$ with common initial and end points,
$$h : [a, b] \times [0, 1] \to U \quad , \quad h(t, s) := (1 - s)\, c_0(t) + s\, c_1(t)$$
is a homotopy. ♢

A.2 Manifolds

In Definition 2.34, we introduced submanifolds of $\mathbb{R}^n$. Here we will define manifolds without reference to such an imbedding.

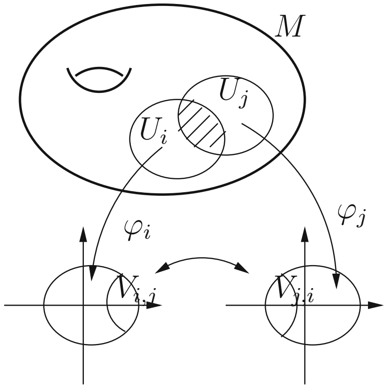
Roughly speaking, an n-dimensional manifold is a topological space that locally looks like $\mathbb{R}^n$. The following definition makes this precise.

A.25 Definition *Let* $(M, \mathcal{O})$ *be a topological space.*

- *M is called **locally Euclidean** if there exists an $n \in \mathbb{N}_0$ such that every $m \in M$ has a neighborhood $U \in \mathcal{O}$ that is homomorphic to $\mathbb{R}^n$. The (unique) number n is then called the **dimension of** M.*
- *If $(M, \mathcal{O})$ is also paracompact (hence in particular Hausdorff), then M is called a **topological manifold**.*
- *A **chart** (U, φ) of $(M, \mathcal{O})$ (also called a **local coordinate system**) consists of an open subset $U \subseteq M$ and a homeomorphism $\varphi : U \to V$ onto its open image $V := \varphi(U) \subseteq \mathbb{R}^n$.*
- *For $r \in \mathbb{N}$ resp. $r = \infty$, two charts (U_i, φ_i), (U_j, φ_j) are said to have a C^r-**overlap** (or are called C^r-**compatible**) if for $V_{i,j} := \varphi_i(U_i \cap U_j) \subseteq \mathbb{R}^n$, the **coordinate change maps***

$$\varphi_{i,j} := \varphi_j \circ \varphi_i^{-1}\restriction_{V_{i,j}} : V_{i,j} \to V_{j,i}$$

are r times continuously differentiable diffeomorphisms (i.e., $\varphi_{i,j} \in C^r\big(V_{i,j}, V_{j,i}\big)$ and $\varphi_{j,i} \in C^r\big(V_{j,i}, V_{i,j}\big)$), see the figure.



Coordinate change map between (U_i, φ_i) and (U_j, φ_j)

- *A C^r-**atlas** on M is a set $\{(U_i, \varphi_i) \mid i \in I\}$ of C^r-compatible charts that cover M, i.e., $M = \cup_{i \in I} U_i$.*
- *Two C^r-atlases are called **equivalent** if any two of their charts are C^r-compatible.*

A.26 Remark

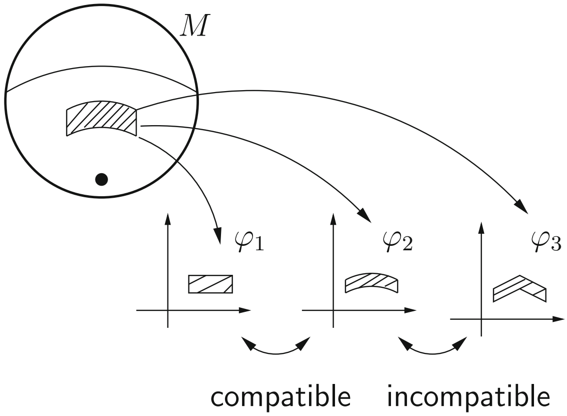
1. Topological manifolds M (or more generally locally Euclidean spaces) have a unique dimension (written: $\dim(M)$). This is not so easy to prove (think of the space-filling Peano curves).
2. Equivalence of C^r-atlases is obviously an equivalence relation. For a C^r-atlas Φ, there exists a unique maximal C^r-atlas Ψ on M that contains Φ. Such a maximal atlas on M is called a C^r**-differentiable structure**. ◊

A.27 Definition *A topological manifold $(M, \mathcal{O})$ together with a C^r-differentiable structure is called a C^r-**manifold**.*

A.28 Remark

1. It is very much intended for the notion *atlas* to be reminiscent of a world atlas in geography. It should show the entire surface of the earth. One map (chart) is not sufficient for this, as is well known. The maps (charts) in an atlas can be obtained by different kinds of projections. An object that looks rectangular in one map may be delimited by curves in another. Creases however are not permissible (see figure).

So we abstract from a metric structure, but do maintain a differentiable structure.



Compatible and incompatible charts

2. The paracompactness requirement (see Definition A.11) is not a limitation in practice. It is automatically fulfilled if the topology of the locally Euclidean space is generated by a metric, for instance if M is defined as a subset of some $\mathbb{R}^m$.
   The *line with two origins* is an example of a one dimensional locally Euclidean topological space that is not Hausdorff (and therefore not paracompact): Here one defines $M := \tilde{M}/\sim$ with $\tilde{M} := \mathbb{R}\times\{0\} \ \cup\ \mathbb{R}\times\{1\}$ and an equivalence relation generated by $(x,0)\sim(x,1)$ for $x\in\mathbb{R}\setminus\{0\}$, and equips M with the quotient topology.
3. We will normally study *smooth* (C^∞-) manifolds M. Then on M, every open cover $(U_i)_{i\in I}$ has a subordinate partition of unity with *smooth* functions $(f_i)_{i\in I}$.
4. One frequently uses the notation M^n for an n-dimensional manifold M. ◊

A.29 Example (Manifolds)

1. Every open subset $M\subset\mathbb{R}^n$ becomes a manifold with the chart (M, Id_M).
2. As a subset of $\mathbb{R}^{n+1}$, the n-sphere $S^n = \{x\in\mathbb{R}^{n+1} \mid \|x\| = 1\}$ is a topological space. The $2n+2$ chart domains

$$U_{\pm j} := \{x\in S^n \mid \pm x_j > 0\} \qquad (j = 1,\ldots,n+1)$$

   and the mappings

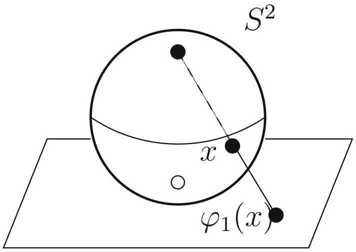
$$\varphi_{\pm j} : U_{\pm j} \longrightarrow \mathbb{R}^n \quad , \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_{n+1}\end{pmatrix} \longmapsto \varphi_{\pm j}(x) = \begin{pmatrix} x_1 \\ \vdots \\ \widehat{x_j} \\ \vdots \\ x_{n+1}\end{pmatrix}$$

make up an atlas on S^n. Here the hat over x_j is to indicate that this coordinate should be omitted. With this atlas, S^n becomes an n-dimensional manifold.

3. A compatible atlas on S^n is given by the two charts $\varphi_k : U_k \to \mathbb{R}^n$ of the *stereographic projection* (see figure), with

$$U_{1/2} := S^n \backslash \{(0, \ldots, 0, \pm 1)^\top\},$$

$$\varphi_{1/2}(x) := \frac{2}{1 \mp x_{n+1}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$



Stereographic projection

As $\frac{1-x_{n+1}}{1+x_{n+1}} = (1-x_{n+1})^2 / \sum_{k=1}^n x_k^2 = 4/\|\varphi_1(x)\|^2$, the coordinate change map on $U_1 \cap U_2$ is, geometrically speaking, the inversion in a sphere of radius 2:

$$\varphi_2 \circ \varphi_1^{-1}(y) = \frac{1 - x_{n+1}}{1 + x_{n+1}} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \frac{4y}{\|y\|^2} \qquad \big(y \in \varphi_1(U_1 \cap U_2) = \mathbb{R}^n \setminus \{0\}\big).$$

In mathematical physics, S^2 occurs as the configuration space of a spherical pendulum, and S^{2n-1} as the energy shell $H^{-1}(\frac{1}{2})$ of the harmonic oscillator

$$H : \mathbb{R}^n_p \times \mathbb{R}^n_q \to \mathbb{R} \quad , \quad H(p, q) := \tfrac{1}{2}\big(\|p\|^2 + \|q\|^2\big).$$

4. In Definition A.27, we completely have disregarded any possible imbedding of the manifold.
We can even define manifolds M by merely giving the images of charts and compatible coordinate change maps. Then the set M is viewed as the set of equivalence classes of points that are equivalent under coordinate change maps, and the topology is the quotient topology.
Here is an example: Whereas the *Möbius strip* was introduced in Example 2.35 as a submanifold of $\mathbb{R}^3$, we can instead define it as an abstract C^∞-manifold M in the following manner: The images of the three charts $\varphi_i : M \to \mathbb{R}^2,\ i = 1, 2, 3$ are the open rectangles $V_i := (0, 3) \times (-1, 1)$, the domains and ranges of the coordinate change maps are the subsets

$$V_{i,i+1} := \{(x, y) \in V_i \mid x > 2\},\ V_{i,i-1} := \{(x, y) \in V_i \mid x < 1\} \quad (i = 1, 2, 3)$$

(where indices are added modulo 3). The coordinate change maps themselves are of the form

$$\varphi_{i,i+1} : V_{i,i+1} \to V_{i+1,i} \quad , \quad (x, y) \mapsto (x - 2, -y)$$

(hence $\varphi_{i+1,i} = \varphi_{i,i+1}^{-1} : V_{i+1,i} \to V_{i,i+1},\ (x, y) \mapsto (x + 2, -y)$).

5. The configuration space of *two* planar pendula is the torus $\mathbb{T}^2 := S^1 \times S^1$. ♢

A.30 Remark (Construction of Manifolds)

1. An *open subset* $N \subseteq M$ of a C^r-manifold $(M, \mathcal{O}_M)$ with atlas $\{(U_i, \varphi_i) \mid i \in I\}$ is again a C^r-manifold in canonical manner.
   As a topology for $(N, \mathcal{O}_N)$, one can take the subspace topology
   $$\mathcal{O}_N := \{U \cap N \mid U \in \mathcal{O}_M \,;$$
   as an *atlas*, the set
   $$\bigl\{\bigl(U_i \cap N, \varphi_i\restriction_N\bigr) \mid i \in I\bigr\}$$
   of the charts (U_i, φ_i) of M restricted to N.
   The case $M = \mathbb{R}^n$ was already discussed in Example A.29.1.
2. Along with differentiable manifolds M and N and their respective atlases $\{(U_i, \varphi_i) \mid i \in I\}$ and $\{(V_j, \psi_j) \mid j \in J\}$, their topological *product space* $M \times N$ is also a differentiable manifold with the atlas
   $$\Bigl\{\bigl(U_i \times V_j, \varphi_i \times \psi_j\bigr) \Bigm| (i, j) \in I \times J\Bigr\}. \qquad \Diamond$$

The closed ball $B^n = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$ is an example for a so-called manifold with boundary. B^n is not a manifold, because the boundary points $x \in \partial B^n = S^{n-1}$ do not have a neighborhood that would be homeomorphic to $\mathbb{R}^n$. There is however a neighborhood that is homeomorphic to a *half space* as is defined by a linear mapping $\ell : \mathbb{R}^n \to \mathbb{R}$, where

$$H^n_\ell := \{y \in \mathbb{R}^n \mid \ell(y) \geq 0\} \qquad \text{(hence with } H^n_0 = \mathbb{R}^n\text{)}.$$

A.31 Definition

- *A paracompact topological space $(M, \mathcal{O})$ for which there exists some $n \in \mathbb{N}_0$ such that every $m \in M$ has a neighborhood $U \in \mathcal{O}$ that is homeomorphic to a half space H^n_ℓ is called a* ***topological manifold with boundary****.*
- *n is then called the* ***dimension of*** *M.*
- *A* ***chart*** *(U, φ) of $(M, \mathcal{O})$ is defined on an open subset $U \subseteq M$, and $\varphi : U \to \varphi(U)$ is a homeomorphism onto the relatively open image $\varphi(U) \subseteq H^n_\ell$.*

C^r**-compatibility** of charts, C^r**-atlases** $\{(U_i, \varphi_i) \mid i \in I\}$ of M, and C^r**-manifolds with boundary** are defined in analogy to Definitions A.25 and A.27.

A.32 Definition
The ***boundary*** *of the manifold with boundary M is the set*

$$\partial M := \bigl\{m \in M \mid \exists\, i \in I : m \in U_i,\ \ \varphi_i(m) \in \partial\bigl(\varphi_i(U_i)\bigr) \subseteq \mathbb{R}^n\bigr\}.$$

The point in this definition is that the boundary of the subset $\varphi_i(U_i)$ of $\mathbb{R}^n$ is defined in the *topological* sense (Definition A.9). It consists exactly of those points of $\varphi_i(U_i)$ that also lie in the hyperplane $\partial H^n_{\ell_i}$ of $\mathbb{R}^n$.

A.33 Theorem
Let M be a C^r-manifold with boundary as in Definition A.32. For the set $J := \{j \in I \mid U_j \cap \partial M \neq \emptyset\}$ of indices and the mappings

$$\tilde{\varphi}_j := \varphi\restriction_{\tilde{U}_j} : \tilde{U}_j \to \partial H^n_{\ell_j} \text{ with the domain } \tilde{U}_j := U_j \cap \partial M\,,$$

$(\tilde{U}_j, \tilde{\varphi}_j)_{j\in J}$ is a C^r-atlas of the boundary of M.

In particular, the boundary ∂M is itself a manifold (without boundary): $\partial\partial M = \emptyset$. In the theory of integration, this is dual to the property $dd = 0$ of the exterior derivative from Theorem B.20.

A.34 Example (manifold with boundary) Half-infinite cylinder

$$M := \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 = R^2,\ x_3 \geq 0\}$$

with radius $R > 0$. We can use, e.g., the following four charts $(U_i^{\pm}, \varphi_i^{\pm})$, $i = 1, 2$ with $U_i^{\pm} := \{x \in M \mid \pm x_i > 0\}$ and $\varphi_i^{\pm} : U_i^{\pm} \to \mathbb{R} \times [0, \infty)$:

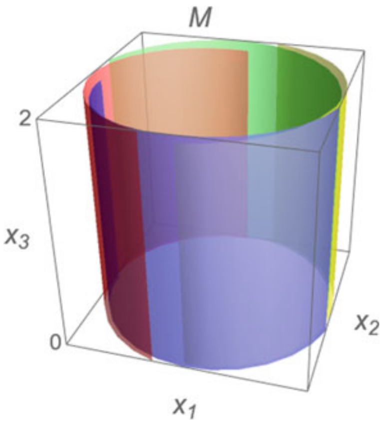
$$\varphi_1^{\pm}(x_1, x_2, x_3) := (\pm x_2, x_3),$$

$$\varphi_2^{\pm}(x_1, x_2, x_3) := (\pm x_1, x_3).$$

The boundary of the cylinder is

$$\partial M = \{x \in M \mid x_3 = 0\},$$

namely a circle with radius R in the (x_1, x_2)-plane of $\mathbb{R}^3$. ◊



A.35 Remark The cartesian product $M \times N$ of two manifolds with boundary is in general not a manifold with boundary. ◊

How can we give and describe a continuous mapping $f : M \to N$ between differentiable manifolds? Clearly we do it again by referring to charts (U, φ) of M at $x \in M$ and (V, ψ) of N at $f(x) \in N$, specifically such charts as are adjusted to the mapping: We need $f(U) \subseteq V$. Then the mapping

$$\psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V) \tag{A.2.1}$$

is defined. It is called the *local representation* of f at x. By continuity of f, we can always find such a local representation, if need be by restricting to a smaller open neighborhood $U' \subseteq U$ of x. (The chart $(U', \varphi\restriction_{U'})$ is compatible with the other charts.)

A.36 Definition

- *$f : M \to N$ is called r* ***times continuously differentiable*** *(in formulas: $f \in C^r(M, N)$) if for all $x \in M$, the local representations of f at x are r times continuously differentiable (see Figure A.2.1).*
- *The mappings $f \in C^\infty(M, N)$ are also called* ***smooth****.*
- *A homeomorphism $f \in C^r(M, N)$ is called C^r-**diffeomorphism** if*

$$f^{-1} \in C^r(N, M).$$

A.37 Literature As a reference on the subject of manifolds, Chapter 1 of ABRAHAM and MARSDEN [AM] can be recommended. Global topological questions are discussed in HIRSCH [Hirs]. ◊

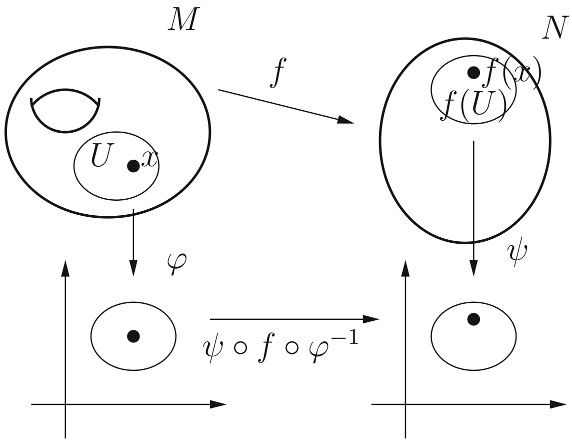
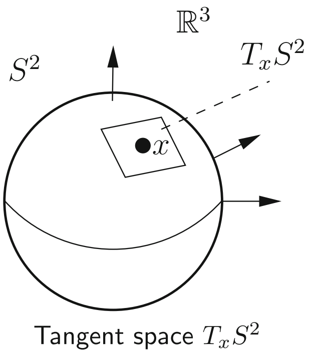


Figure A.2.1 Differentiability of $f : M \to N$

A.3 The Tangent Bundle

What kind of geometric structure is formed by the states (namely positions and velocities) of a mechanical system, given that its configuration space is a manifold M ? They form what is known as the tangent bundle TM of M.



If M is a manifold imbedded into $\mathbb{R}^n$, then it is clear how to understand the tangential space of M at a point $x \in M$. It will be the subspace $T_x M$ of those vectors of the tangential space $T_x\mathbb{R}^n \cong \mathbb{R}^n$ of $\mathbb{R}^n$ at x that are tangential to M. In particular if $M \subset \mathbb{R}^n$ is open, then

$$TM \cong M \times \mathbb{R}^n . \qquad \text{(A.3.1)}$$

A.38 Example The tangent space of the sphere $S^d \subset \mathbb{R}^{d+1}$ at $x \in S^d$ is $T_x S^d = \{y \in \mathbb{R}^{d+1} \mid \langle y, x \rangle = 0\}$, see the figure to the right.[5] ◊

As not all manifolds are defined as subsets of some $\mathbb{R}^n$, we need to approach the general definition of the tangent bundle differently:

A.39 Definition

- *A **tangent vector** of a manifold M at a point $x \in M$ is an equivalence class of curves $c \in C^1\big(]-\varepsilon, \varepsilon[, M\big)$ with $c(0) = x$, where two such curves c_1, c_2 are deemed **equivalent** if in some chart (U, φ), $x \in U$, one has*

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi \circ c_1(t)\Big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\varphi \circ c_2(t)\Big|_{t=0}.$$

- *The set $T_x M$ of tangent vectors of M at x is called the **tangent space of** M at x.*
- *The **tangent bundle** TM of M is the union $\bigcup_{x \in M} T_x M$.*
- *We denote the projection of the tangent vectors in $T_x M$ to their **foot** or **base point** x as $\pi_M : TM \to M$; then $\pi_M^{-1}(x) = T_x M$ is called **fiber** above $x \in M$.*
- *A continuous mapping $v : M \to TM$ with $\pi_M \circ v = Id_M$ is called a **vector field** on M, see Figure A.3.2.*

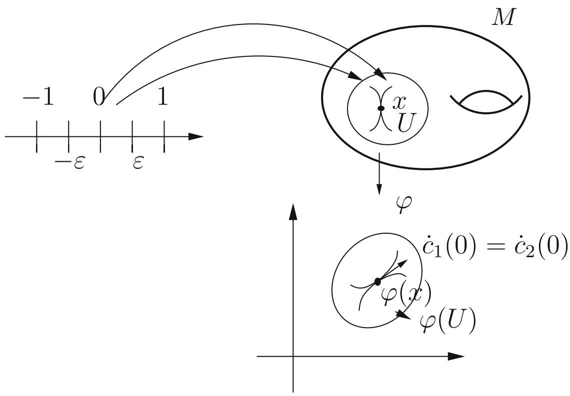


Figure A.3.1 Tangent vector as an equivalence class of curves

So the tangent vector is defined as the set of all those curves that are tangential to each other in the sense of

$$\varphi\big(c_1(t)\big) - \varphi\big(c_2(t)\big) = o(t)\,, \tag{A.3.2}$$

[5]Example A.38 shows that, already by reasons of dimension, the tangent bundle *TM* of a submanifold $M \subset \mathbb{R}^d$ can in general *not* be imbedded into $\mathbb{R}^d$; the tangent spaces $T_x M$ however can be so imbedded.

see Figure A.3.1. Whereas the property of tangentiality (A.3.2) of two curves is defined in a chart, it is invariant under coordinate changes.

A.40 Remark (Definitions of the Tangent Bundle)
Let $M \subseteq \mathbb{R}^n$ be open, and $(U, \varphi) := (M, \mathrm{Id})$ be used as a chart. Then Definition A.39 yields, in accordance with (A.3.1), that $TM \cong M \times \mathbb{R}^n$; consequently the tangent bundle of M is a differentiable manifold of twice the dimension.

In the general setting, the definition of a tangent bundle TM of a submanifold $M \subseteq \mathbb{R}^n$ as given at the beginning of this section is transformed into Definition A.39 of TM if we assign to a tangent vector $v \in T_m\mathbb{R}^n$ at $m \in M$ the equivalence class of the curve $t \mapsto m + t\ v$ projected onto M. ♢

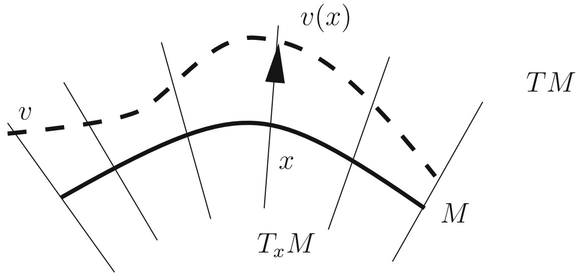


Figure A.3.2 A tangent vector field $v : M \to TM$; the zero section of TM is identified with M itself

In the general setting, for a point x on a manifold M, a chart (U, φ) at x and a C^1-curve $c : (-\varepsilon, \varepsilon) \to U$ with $c(0) = x$, the time derivatives

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi \circ c(t)\Big|_{t=0}$$

are vectors in $T_{\varphi(x)}\mathbb{R}^n \cong \mathbb{R}^n$, and in the image of the chart, we can multiply these vectors by a number or add them to each other. This structure of a vector space carries over to the set T_xM of tangent vectors to M in x in a way that is independent of the coordinate chart.

A.41 Definition *For $f \in C^1(M, N)$, the mapping $Tf : TM \to TN$ with*

$$Tf\big([c]_x\big) := [f \circ c]_{f(x)} \qquad \big(x \in M,\ c \text{ curve at } x\big)$$

*is called the **tangential mapping** of f (where $[\cdot]$ denotes the equivalence class).*

A.42 Theorem

- *For a C^{r+1}-manifold M, the tangent space T_xM of M at x is a real vector space of dimension* $\dim(T_xM) = \dim(M)$.

- *The tangent bundle TM is a C^r-manifold, and*

$$\dim(TM) = 2\,\dim(M)\,.$$

Proof: Let $\mathcal{A} := \{(U_i, \varphi_i) \mid i \in I\}$ be an atlas of M. Then

$$T\mathcal{A} := \{(TU_i, T\varphi_i) \mid i \in I\}$$

is an atlas of *TM*, called the *natural atlas* . Its charts are called *natural charts.* □

While in principle, we can use arbitrary coordinates in the manifold *TM*, it is sensible to use linear coordinates for the tangent vectors, so that tangent vectors at a point can be added in the charts as usual. A chart (U, φ) on M induces on U the $n = \dim(M)$ vector fields

$$\frac{\partial}{\partial\varphi_1}, \ldots, \frac{\partial}{\partial\varphi_n} : U \to TU\,,$$

which are mapped under the tangential mapping into

$$T\varphi\left(\frac{\partial}{\partial\varphi_l}\right)(u) = \big(\varphi(u), e_l\big) \qquad (u \in U,\ l = 1, \ldots, n) \tag{A.3.3}$$

(with e_l denoting the l^{th} canonical basis vector of $\mathbb{R}^n$). For $x \in U$, the tangent vectors $\frac{\partial}{\partial\varphi_1}(x), \ldots, \frac{\partial}{\partial\varphi_n}(x)$ are a basis of T_xM.

The set $\mathcal{X}(M)$ of vector fields on a manifold M is an $\mathbb{R}$-vector space. Within the domain U of a chart, $X \in \mathcal{X}(M)$ has the unique representation $X(x) = \sum_{k=1}^n X_k(x)\frac{\partial}{\partial\varphi_k}(x)$ with continuous functions $X_k : U \to \mathbb{R}$.

A.43 Definition

- *The tangent bundle TM of an n-dimensional manifold M is called* ***parallelizable*** *if there exists a diffeomorphism*

$$I : TM \to M \times \mathbb{R}^n$$

that is linear in the fibers (i.e., linear if restricted to each of the fibers T_mM for any $m \in M$) and is the identity with respect to M, i.e.,

$$I \circ \pi_M^{-1}(m) = \{m\} \times \mathbb{R}^n\,.$$

- *In this case, I is called a* ***parallelization*** *of TM.*

All parallelizable manifolds are in particular, see Definition F.12.

A.44 Example (Parallelizability)

1. Lie groups G are parallelizable, because with the left operation L_g from (E.1.3) and the Lie algebra $\mathfrak{g} \cong T_eG \cong \mathbb{R}^{\dim(G)}$ of G, the mapping

$$G \times \mathfrak{g} \to TG \quad , \quad (g, \xi) \mapsto (T_e L_g)(\xi)$$

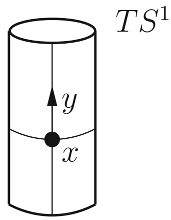
is a diffeomorphism that is linear on the fibers satisfying $(T_e L_g)(\xi) \in T_g G$.

2. The tangent bundle of the sphere $S^n = \{x \in \mathbb{R}^{n+1} \mid \|x\| = 1\}$ is

$$TS^n = \left\{(x, y) \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \mid \|x\| = 1 \, , \, \langle x, y \rangle = 0\right\}.$$

• $\boldsymbol{TS^1}$: Since $S^1 = \{x \in \mathbb{C} \mid |x| = 1\}$, we can identify the tangent bundle with

$$TS^1 = \left\{(x, y) \in \mathbb{C} \times \mathbb{C} \mid |x| = 1 \, , \, \frac{y}{x} \in \imath\,\mathbb{R}\right\}.$$



Tangent space of the circle S^1

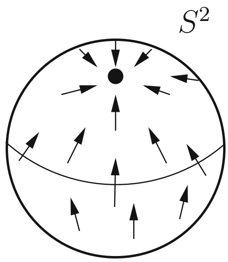
We find a parallelization

$$I : TS^1 \to S^1 \times \mathbb{R} \quad , \quad (x, y) \mapsto \big(x, y/(\imath\, x)\big)$$

of TS^1, with the inverse $I^{-1}(x, z) = (x, \imath\, xz)$, see the figure to the right. We will take advantage of this fact when studying the planar pendulum.

• $\boldsymbol{TS^2}$**. Claim:** TS^2 ist *not* parallelizable.[6]

Proof: Indirect. Consider $I^{-1}\left(\left\{x \times \binom{1}{0}\right\}\right)$. This would have to be a tangent vector at $x \in S^2$. This tangent vector is not zero by hypothesis (linearity). Considering these tangent vectors for all $x \in S^2$, we obtain a non-vanishing vector field on S^2. However, such a vector field $Y : S^2 \to TS^2$ does not exist (see figure). Because otherwise let $Y(x)$ be of length 1 (as could be achieved by normalization) for all $x \in S^2$ hence $Y_\varepsilon := \varepsilon\, Y$ of length $|\varepsilon|$. Then



tangent vector field on S^2

$$f_\varepsilon : S^2 \to \mathbb{R}^3 \quad , \quad x \mapsto x + Y_\varepsilon(x)$$

maps onto the sphere $S^2(r)$ of radius $r := \sqrt{1 + \varepsilon^2}$, and f_ε is, for small $|\varepsilon|$, a diffeomorphism of the spheres.

Now consider on $\mathbb{R}^3$ the 2-form (see Appendix B.2)

$$\omega := x_1\, dx_2 \wedge dx_3 + x_2\, dx_3 \wedge dx_1 + x_3\, dx_1 \wedge dx_2 = r^3 \cos(\theta)\, d\varphi \wedge d\theta$$

in spherical coordinates

$$x_1 = r \cos(\theta) \cos(\varphi) \quad , \quad x_2 = r \cos(\theta) \sin(\varphi) \quad , \quad x_3 = r \sin(\theta) \, .$$

---

[6]**Hairy Ball Theorem**: Every continuously combed hairy ball has at least one bald spot.
The following proof by Milnor generalizes to all spheres S^{2n}, see GALLOT, HULIN and LAFONTAINE [GHL], Chapter I.C.
Among the odd-dimensional spheres, the only parallelizable ones are, apart from S^1, only S^3 and S^7, see HIRZEBRUCH [Hirz]. The fact that S^3 is parallelizable, can be seen from the fact that S^3 is diffeomorphic to the Lie Group SU(2) (see (E.2.2)).

Now we can calculate the surface area $F(r)$ of the sphere $S^2(r)$ on one hand by

$$F(r) = \frac{1}{r}\int_{S^2(r)} \omega = 4\pi r^2 = 4\pi\,(1+\varepsilon^2)\,,$$

on the other hand by our hypothesis for contradiction as

$$F(r) = \frac{1}{r}\int_{f_\varepsilon(S^2(1))} \omega = \frac{1}{r}\int_{S^2(1)} f_\varepsilon^*(\omega)\,.$$

The latter expression however is a polynomial in ϵ divided by $r = \sqrt{1+\varepsilon^2}$, as can be seen by direct calculation of $f_\varepsilon^*(\omega)$. Contradiction! □

The tangent bundle *TM* of a configuration manifold M is the space of all positions and velocities. The Lagrangian of a system in mechanics with configuration space M is thus a function $L : TM \to \mathbb{R}$.

A.45 Definition *For C^1-manifolds M and N, consider $f \in C^1(M, N)$.*

- *f is called **immersive at** $m \in M$ if $T_m f : T_m M \to T_{f(m)} N$ is injective; f is called **submersive at** $m \in M$, and m a **regular point** of f, if $T_m f$ is surjective. Otherwise, m is called a **singular point** of f.*
- *f is called an **immersion** if for all $m \in M$, f is immersive at m. f is called a **submersion** if for all $m \in M$, f is submersive at m.*
- *f is called an **imbedding** if f is an immersion that maps M homeomorphically onto $f(M)$. (In symbols: $f : M \hookrightarrow N$)*
- *$n \in N$ is called a **regular value** of f if all $m \in f^{-1}(n)$ are regular points; otherwise a **singular value**.*

Immersions need not be injective, regular values $n \in N$ need not be in the image $f(M)$.

The inverse mapping theorem implies:

A.46 Theorem (Regular Value Theorem)
For $r \geq 1$ and C^r-manifolds M and N, let $n \in N$ be a regular value of $f \in C^r(M, N)$.
Then $U := f^{-1}(n) \subseteq M$ is a C^r-submanifold, and $\dim U = \dim M - \dim N$.

Many phenomena can already be observed in curves $c : I \to N$ in a manifold N. Such a C^1-curve is called *regular* if it is an immersion, i.e., if the velocity vector $c'(t) \in T_{c(t)} M$ never vanishes.

A.47 Exercises (**Differential Topology**) Show:

1. $f : \mathbb{R} \to \mathbb{R}\,, t \mapsto t^3$ is injective, but is not immersive at $t = 0$. The image is $f(\mathbb{R}) = \mathbb{R}$, a manifold after all.

2. $f : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto \binom{t^3}{t^2}$ (see Figure A.3.3, left) is a smooth mapping, but is not immersive at $t = 0$, and the image is not a submanifold (compare with the implicitly defined case in Example 2.42).
3. The curves $f_k : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto \cos(kt) \binom{\sin t}{\cos t}$ are called *rose* or *rhodonea curves* (see Figure A.3.3, center). The f_k are immersions for all $k \in \mathbb{R}$, but in general the images $f_k(\mathbb{R}) \subset \mathbb{R}^2$ are not submanifolds.
4. $f : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto \binom{\sin t}{\cos t}$ is a non-injective immersion, hence not an imbedding. Nevertheless, the image $S^1 \subset \mathbb{R}^2$ is a submanifold.
   In contrast, $\tilde{f} : \mathbb{R}/(2\pi\mathbb{Z}) \to \mathbb{R}^2$, $t + 2\pi\mathbb{Z} \mapsto \binom{\sin t}{\cos t}$ is well-defined and now imbeds the one dimensional manifold $\mathbb{R}/(2\pi\mathbb{Z})$.
5. $f : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto \exp(-t^2) \binom{t}{t^3}$ (see Figure A.3.3, right) is an injective immersion, but not an imbedding.

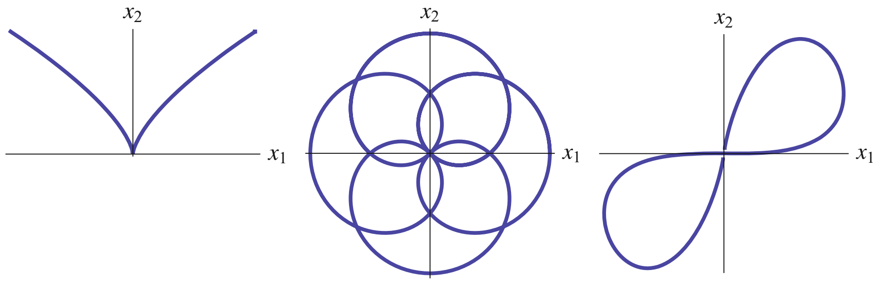


Figure A.3.3 Left: Image of a smooth, non-immersive mapping. Center: Rose petal curve with $k = 2/3$, the image of a non-injective immersion. Right: Image of an injective immersion that is not an imbedding

6. $f \in C^1(M, N)$ cannot be an immersion if $\dim(M) > \dim(N)$, and cannot be a submersion if $\dim(M) < \dim(N)$.
7. The projections $\pi : E \to B$ of differentiable fiber bundles, which are discussed in Appendix F.1, are submersions. ◊

In differential topology, there is a close connection between statements about manifolds and statements about mappings. The following theorem may serve as an example:

A.48 Theorem
Let N be a C^r-manifold, where $r \geq 1$. A subset $A \subset N$ is a C^r-submanifold if and only if A is the image of a C^r-imbedding of some manifold.

On the other hand, all abstractly defined manifolds can be viewed as submanifolds of some $\mathbb{R}^d$ according to the following theorem:

A.49 Theorem (Whitney Imbedding Theorem)
Every compact n-dimensional differentiable manifold can be imbedded into $\mathbb{R}^{2n}$.

A.50 Remark

This bound, linear in the dimension, is sharp, as can be seen in Example A.47.4 of S^1 or the real projective space $\mathbb{R}\mathrm{P}(2)$, which cannot be imbedded into $\mathbb{R}^3$ as it is a compact, non-orientable surface. There do however exist immersions $\mathbb{R}\mathrm{P}(2) \to \mathbb{R}^3$, e.g., *Boy's surface*, which serves as an emblem for the Oberwolfach Research Institute for Mathematics (Mathematisches Forschungsinstitut Oberwolfach).[7] ◊

Riemannian Manifolds

A.51 Definition *Let M be a differentiable manifold.*

- *A* ***Riemannian metric*** *on M is a family $g = (g_m)_{m\in M}$ of positive definite symmetric bilinear forms*
$$g_m : T_m M \times T_m M \to \mathbb{R} \qquad (m \in M)$$
that depend differentiably on m.
- *For a Riemannian metric g on M, the pair (M, g) is called a* ***Riemannian manifold****.*

The function g is also called *metric tensor*.[8] From the example $(\mathbb{R}^n, g)$ with the translation invariant Riemannian metric $g_m(v, w) := \langle v, w\rangle$, one obtains again Riemannian manifolds by restriction of g to submanifolds $M \subseteq \mathbb{R}^d$, see page 517.

As a non-degenerate bilinear form, the Riemannian metric defines an isomorphism between tangent and cotangent spaces, namely

$$T_m M \to T_m^* M \quad , \quad v \mapsto g_m(v, \cdot) \qquad (m \in M).$$

This gives rise to an isomorphism of vector bundles, $\flat : TM \to T^*M$. This Legendre transform is called *musical isomorphism*, and its inverse is denoted as $\natural : T^*M \to TM$.

The *gradient* $\nabla f : M \to TM$ of a function $f \in C^1(M, \mathbb{R})$ is the vector field that arises under $\natural$ from the exterior derivative $df : M \to T^*M$. Therefore, ∇f depends on the metric tensor g, in contradistinction to df.

A.52 Literature Well-known books on differential topology are [Hirs] by Hirsch and [BJ] by Bröcker and Jänich. The two volume opus [CDD] by Choquet-Bruhat, DeWitt- Morette and Dillard- Bleick gives an extensive survey of the theory of differentiable manifolds, including differential forms, theory of bundles, and differential geometry. ◊

[7] Image: courtesy of Oberwolfach Research Institute for Mathematics.

[8] It is not a metric in the sense of metric spaces; however it gives rise to the definition of such a metric, see (G.3.3).

Appendix B
Differential Forms

In numerous applications of analysis to physics, one gets to *integrate over submanifolds* of $\mathbb{R}^n$, for instance when determining

– the magnetic flux through a surface that is bounded by a loop of electric conductor
– the work performed along a path, etc.

To perform such integrals, Élie Cartan and others developed the calculus of differential forms on manifolds. This calculus also highlights the *geometric content* of theories in physics like classical mechanics, electrodynamics, or general relativity; for instance Maxwell's equations can be written in terms of differential forms as $dF = 0$, $\delta F = j$, see Example B.21.

B.1 Literature A good introduction can be found in the book [AF] by Agricola and Friedrich. ◊

The first step is the algebraic theory of *exterior forms*, because it is them that describe the local behavior of differential forms at one point of the manifold.

B.1 Exterior Forms

B.2 Definition *Let E be an n-dimensional real vector space. A mapping $\varphi : E \times \ldots \times E \to \mathbb{R}$ is called* ***multilinear*** *if it is linear in each of its arguments, i.e., for $\lambda \in \mathbb{R}$, $j = 1, \ldots, k$, and x_j, x_j^I, $x_j^{II} \in E$,*

$$\varphi(x_1, \ldots, x_{j-1}, \lambda x_j, x_{j+1}, \ldots, x_k) = \lambda\varphi(x_1, \ldots, x_k)$$

and

$$\begin{aligned}&\varphi\big(x_1, \ldots, x_{j-1}, x_j^I + x_j^{II}, x_{j+1}, \ldots, x_k\big)\\ &\quad = \varphi\big(x_1, \ldots, x_j^I, \ldots, x_k\big) + \varphi\big(x_1, \ldots, x_j^{II}, \ldots, x_k\big)\,.\end{aligned}$$

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

*More precisely, one refers to such a φ as a k-**linear** mapping.*

On $E := \mathbb{R}^n$ with the standard basis $e_1, \dots, e_n \in E$, let $\alpha_1, \dots, \alpha_n \in E^*$ denote the *dual basis* (which means that $\alpha_i(e_j) = \delta_{ij}$).

B.3 Example (Exterior Forms)

1. $k = 1$. Then φ is a *linear form* on E, and for $\varphi \neq 0$, the pre-image $\varphi^{-1}(0) \subset E$ is a subspace of dimension $n - 1$.
2. $k = 2$, $E = \mathbb{R}^n$ with the inner product $\langle \cdot, \cdot \rangle$.
   For $A \in \mathrm{Mat}(n, \mathbb{R})$, a *bilinear form* is given by $\varphi : E \times E \to \mathbb{R}$, $\varphi(x, y) := \langle x, Ay \rangle$. It is called *(anti)symmetric*, if $\varphi(x, y) = \pm\varphi(y, x) \quad (x, y \in E)$.
3. $k = n$, $E = \mathbb{R}^n$. Then $\varphi(x_1, \dots, x_n) := \det(x_1, \dots, x_n) \quad (x_1, \dots, x_n \in E)$ defines the *determinant form*, which gives the oriented volume of the parallelotope spanned by the vectors $x_1, \dots, x_n$. ◊

Obviously, two k-linear mappings φ_1, φ_2 can be *added* by setting

$$(\varphi_1 + \varphi_2)(x_1, \dots, x_k) := \varphi_1(x_1, \dots, x_k) + \varphi_2(x_1, \dots, x_k) \qquad (x_1, \dots, x_k \in E), \tag{B.1.1}$$

and likewise, we can *multiply* a k-linear mapping φ with a real number by

$$(\lambda\varphi)(x_1, \dots, x_k) := \lambda\big(\varphi(x_1, \dots, x_k)\big) \qquad (\lambda \in \mathbb{R},\ x_1, \dots, x_k \in E). \tag{B.1.2}$$

This makes the set $L^k(E, \mathbb{R})$ of k-linear mappings to $\mathbb{R}$ into an $\mathbb{R}$-*vector space*.

B.4 Definition *Let E be an n-dimensional $\mathbb{R}$-vector space.*
*Then $\varphi \in L^k(E, \mathbb{R})$ is called an **exterior k-form** if it is **antisymmetric**, i.e.,*

$$\varphi(x_1, \dots, x_i, \dots, x_j, \dots, x_k) = -\varphi(x_1, \dots, x_j, \dots, x_i, \dots, x_k) \qquad (x_\ell \in E).$$

The subspace of exterior k-forms is denoted as $\mathbf{\Omega}^k(E) \subset L^k(E, \mathbb{R})$.

B.5 Example (Spaces of Exterior Forms)

1. $\mathbf{\Omega}^1(E) = L^1(E) \cong E^*$.
2. The bilinear mapping $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $(x, y) \mapsto \langle x, Ay \rangle$ defines an exterior 2-form on $\mathbb{R}^n$ if and only if the matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ is antisymmetric, i.e., $A^\top = -A$. So we conclude $\dim(\mathbf{\Omega}^2(\mathbb{R}^n)) = \binom{n}{2}$.
3. The exterior n-forms on $\mathbb{R}^n$ are the multiples of the determinant form. ◊

B.6 Definition *The **exterior product** of $\omega_1, \dots, \omega_k \in \mathbf{\Omega}^1(E)$ is defined by*

$$\omega_1 \wedge \dots \wedge \omega_k(x_1, \dots, x_k) := \det \begin{pmatrix} \omega_1(x_1) & \dots & \omega_k(x_1) \\ \vdots & & \vdots \\ \omega_1(x_k) & \dots & \omega_k(x_k) \end{pmatrix} \quad (x_1, \dots, x_k \in E).$$

Clearly $\omega_1 \wedge \ldots \wedge \omega_k$ is a k-form, i.e., is in $\mathbf{\Omega}^k(E)$. In particular,

$$\alpha_{i_1} \wedge \ldots \wedge \alpha_{i_k} \in \mathbf{\Omega}^k(E) .$$

This exterior form coincides up to a sign with the one in which the indices $i_1, \ldots, i_k$ are arranged in increasing order, and it is nonzero exactly if all the indices are distinct.

Now every k-form $\omega \in \mathbf{\Omega}^k(\mathbb{R}^n)$ can uniquely be written as a linear combination

$$\omega = \sum_{1 \le i_1 < \ldots < i_k \le n} \omega_{i_1 \ldots i_k} \; \alpha_{i_1} \wedge \ldots \wedge \alpha_{i_k}$$

with coefficients $\omega_{i_1 \ldots i_k} := \omega(e_{i_1}, \ldots, e_{i_k}) \in \mathbb{R}$. For $\dim(E) = n$, one concludes therefore that

$$\dim\big(\mathbf{\Omega}^k(E)\big) = \binom{n}{k}.$$

The *exterior product*, or *wedge product* of the k-form ω with an l-form

$$\psi = \sum_{1 \le j_1 < \ldots < j_l \le n} \psi_{j_1 \ldots j_l} \; \alpha_{j_1} \wedge \ldots \wedge \alpha_{j_l} \;\in\; \mathbf{\Omega}^l(E)$$

is now defined as

$$\omega \wedge \psi := \sum_{1 \le i_1 < \ldots < i_k \le n} \; \sum_{1 \le j_1 < \ldots < j_l \le n} \omega_{i_1 \ldots i_k} \psi_{j_1 \ldots j_l} \; \alpha_{i_1} \wedge \ldots \wedge \alpha_{i_k} \wedge \alpha_{j_1} \wedge \ldots \wedge \alpha_{j_l} \, ,$$

compatible with the distributive law. All those terms in the sum that have some $i_r = j_s$, will be 0, because $\alpha_l \wedge \alpha_l = -\alpha_l \wedge \alpha_l = 0$. The remaining terms in the sum contribute to the exterior form $\omega \wedge \psi \in \mathbf{\Omega}^{k+l}(\mathbb{R}^n)$.

Clearly the exterior product is *associative*, i.e.,

$$(\omega \wedge \psi) \wedge \rho = \omega \wedge (\psi \wedge \rho) .$$

Moreover, for a k-form ω and an l-form ψ, one has

$$\omega \wedge \psi = (-1)^{k\,l} \; \psi \wedge \omega \, ,$$

because we need to commute 1-forms $(k\,l)$ times to obtain one form from the other.

B.7 Definition

- *A k-form $\omega \in \mathbf{\Omega}^k(E)$ is called* ***decomposable*** *if there exist one-forms $\omega_1, \ldots, \omega_k \in \mathbf{\Omega}^1(E)$ such that $\omega = \omega_1 \wedge \ldots \wedge \omega_k$.*
- *The* ***rank*** *of a k-form $\omega \in \mathbf{\Omega}^k(E)$ is the smallest number $n \in \mathbb{N}$ for which we can write $\omega = \sum_{\ell=1}^n \omega^{(\ell)}$ with decomposable $\omega^{(\ell)} \in \mathbf{\Omega}^k(E)$.*

B.8 Example (Symplectic form on $\mathbb{R}^{2n}$)
$\omega := \sum_{i=1}^n \alpha_i \wedge \alpha_{i+n} \in \mathbf{\Omega}^2(\mathbb{R}^{2n})$. The symplectic form ω plays a key role in classical mechanics, see Chapter 6.2. There we call the coordinates $x_1, \ldots, x_n$ momentum coordinates, and the coordinates $x_{n+1}, \ldots, x_{2n}$ position coordinates.

For $n = 2$, one has $\omega = \alpha_1 \wedge \alpha_3 + \alpha_2 \wedge \alpha_4$, hence

$$\begin{aligned}\omega \wedge \omega &= (\alpha_1 \wedge \alpha_3 + \alpha_2 \wedge \alpha_4) \wedge (\alpha_1 \wedge \alpha_3 + \alpha_2 \wedge \alpha_4) \\ &= \underbrace{\alpha_1 \wedge \alpha_3 \wedge \alpha_1 \wedge \alpha_3}_{0} + \alpha_2 \wedge \alpha_4 \wedge \alpha_1 \wedge \alpha_3 + \alpha_1 \wedge \alpha_3 \wedge \alpha_2 \wedge \alpha_4 + \underbrace{\alpha_2 \wedge \alpha_4 \wedge \alpha_2 \wedge \alpha_4}_{0} \\ &= (-1)^3 \alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 + (-1)^1 \alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 = -2\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 .\end{aligned} \tag{B.1.3}$$

B.9 Exercises (Volume Form) Show, generalizing (B.1.3), that the n-fold exterior product $\omega^{\wedge n}$ of ω with itself yields a volume form, more precisely:

$$\bigwedge_{k=1}^{2n} \alpha_k = \frac{(-1)^{\binom{n}{2}}}{n!}\, \omega^{\wedge n}$$

(This implies that the rank of ω equals n). ◊

B.10 Example (Cross Product) We consider a vector

$$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1 e_1 + v_2 e_2 + v_3 e_3 \in \mathbb{R}^3 .$$

- Now, using the canonical inner product $\langle \cdot, \cdot \rangle$ of $\mathbb{R}^3$ for the first time, we assign to the vector v the 1-form $v^* \in \mathbf{\Omega}^1(\mathbb{R}^3)$ by

$$v^*(w) := \langle v, w \rangle = v_1 w_1 + v_2 w_2 + v_3 w_3 \qquad \big(w \in \mathbb{R}^3\big).$$

- Similarly, we can assign to v the 2-form $\omega_v \in \mathbf{\Omega}^2(\mathbb{R}^3)$ by

$$\omega_v(x, y) := \det(v, x, y) \qquad \big(x, y \in \mathbb{R}^3\big).$$

  We find

$$v^* = v_1\alpha_1 + v_2\alpha_2 + v_3\alpha_3 \quad \text{and} \quad \omega_v = v_1 \alpha_2 \wedge \alpha_3 + v_2 \alpha_3 \wedge \alpha_1 + v_3 \alpha_1 \wedge \alpha_2 .$$

- The exterior product of two 1-forms obtained in this way is

$$\begin{aligned} v^* \wedge w^* &= (v_1\alpha_1 + v_2\alpha_2 + v_3\alpha_3) \wedge (w_1\alpha_1 + w_2\alpha_2 + w_3\alpha_3) \\ &= (v_1 w_2 - v_2 w_1)\alpha_1 \wedge \alpha_2 + (v_2 w_3 - v_3 w_2)\alpha_2 \wedge \alpha_3 \\ &\quad + (v_3 w_1 - v_1 w_3)\alpha_3 \wedge \alpha_1 = \omega_{v \times w} .\end{aligned}$$

This way, we have represented the *cross product* of two vectors in $\mathbb{R}^3$ in terms of forms. ◊

B.11 Theorem
The vectors $\omega_1, \ldots, \omega_k \in E^*$ are linearly dependent if and only if

$$\omega_1 \wedge \ldots \wedge \omega_k = 0\,.$$

Proof: Let $\omega_1, \ldots, \omega_k \in E^*$.

- If they are linearly dependent, we can find one index $i \in \{1, \ldots, k\}$ and numbers $c_l \in \mathbb{R}$ with $\omega_i = \sum_{l\in\{1,\ldots,k\}\setminus\{i\}} c_l\, \omega_l$. But then one has, with ω_l in the i^{th} position,

$$\omega_1 \wedge \ldots \wedge \omega_k = \textstyle\sum_{l\in\{1,\ldots,k\}\setminus\{i\}} c_l\; \omega_1 \wedge \ldots \wedge \omega_l \wedge \ldots \wedge \omega_k = 0\,,$$

  because in each term in the sum, ω_l occurs twice.
- Otherwise, we can extend the family of vectors $\omega_1, \ldots, \omega_k$ to a basis

$$\omega_1, \ldots, \omega_n \quad \text{with} \quad n := \dim(E^*)$$

  such that $\omega_1 \wedge \ldots \wedge \omega_n \neq 0$, hence also $\omega_1 \wedge \ldots \wedge \omega_k \neq 0$. □

B.12 Definition *The real vector space*

$$\mathbf{\Omega}^*(E) \;:=\; \bigoplus_{k=0}^{\infty} \mathbf{\Omega}^k(E) \;\cong\; \bigoplus_{k=0}^{\dim(E)} \mathbf{\Omega}^k(E)$$

*(with $\mathbf{\Omega}^0(E) := \mathbb{R}$) with multiplication given by the wedge product is called the **exterior**, or **Grassmann algebra** over E.*

B.13 Remark

1. $\dim\big(\mathbf{\Omega}^*(E)\big) = 2^{\dim(E)}$, because $\sum_{k=0}^{n}\binom{n}{k} = 2^n$.
2. For arbitrary $k, l \in \mathbb{N}_0$ and $\omega \in \mathbf{\Omega}^k(E)$, $\varphi \in \mathbf{\Omega}^l(E)$, one has $\omega \wedge \varphi \in \mathbf{\Omega}^{k+l}(E)$;
3. however for $m > \dim(E)$, one has $\dim\big(\mathbf{\Omega}^m(E)\big) = 0$.
4. Exterior forms have the decomposition $\omega = \oplus_k \omega_k \in \mathbf{\Omega}^*(E) = \oplus_{k=0}^{\dim(E)} \mathbf{\Omega}^k(E)$. In most applications, 'mixed' exterior forms where more than one term ω_k is nonzero, do not occur; in any case, it suffices in the sequel to consider the case of forms in $\mathbf{\Omega}^k$, and then extend linearly to $\mathbf{\Omega}^*$. ◊

B.14 Definition *For a linear mapping $f : E \to F$ of finite dimensional $\mathbb{R}$-vector spaces, $k \in \mathbb{N}_0$, and $\omega \in \mathbf{\Omega}^k(F)$, the k-form $f^*(\omega)$ defined by*

$$f^*(\omega)(v_1, \ldots, v_k) := \omega\big(f(v_1), \ldots, f(v_k)\big) \qquad (v_1, \ldots, v_k \in E)$$

*is called the **pull-back** of ω by f.*

$f^*(\omega)$ is k-linear and antisymmetric, hence $f^*(\omega) \in \mathbf{\Omega}^k(E)$.

B.15 Theorem (pull-back) *Let $f \in Lin(E, F)$.*

1. *The mappings* $f^* : \boldsymbol{\Omega}^k(F) \to \boldsymbol{\Omega}^k(E) \quad (k \in \mathbb{N}_0)$ *are linear.*
2. *For* $g \in Lin(F, G)$, *and thus* $g \circ f \in Lin(E, G)$, *one has* $(g \circ f)^* = f^* \circ g^*$.
3. *If* $f = Id_E$, *then* $f^* = Id_{\boldsymbol{\Omega}^*(E)}$.
4. *If* $f \in GL(E, F)$, *then* $(f^*)^{-1} = (f^{-1})^*$.
5. *For* $\alpha, \beta \in \boldsymbol{\Omega}^*(F)$, *one has* $f^*(\alpha \wedge \beta) = f^*(\alpha) \wedge f^*(\beta)$.

Proof: In the proof, we will assume $v_1, \dots, v_k \in E$ and $\lambda \in \mathbb{R}$.

1. With the vector space structure on $\boldsymbol{\Omega}^k$ explained in (B.1.1) and (B.1.2), one has for $\alpha, \beta \in \boldsymbol{\Omega}^k(F)$:

$$\begin{aligned} f^*(\alpha + \lambda\beta)(v_1, \dots, v_k) &= (\alpha + \lambda\beta)\big(f(v_1), \dots, f(v_k)\big) \\ &= \alpha\big(f(v_1), \dots, f(v_k)\big) + \lambda\beta\big(f(v_1), \dots, f(v_k)\big) \\ &= \big(f^*\alpha + \lambda f^*\beta\big)\big(v_1, \dots, v_k\big)\,. \end{aligned}$$

2. For $\alpha \in \boldsymbol{\Omega}^k(G)$, one has

$$\begin{aligned} (g \circ f)^*\alpha(v_1, \dots, v_k) &= \alpha\big(g \circ f(v_1), \dots, g \circ f(v_k)\big) \\ &= g^*\alpha\big(f(v_1), \dots, f(v_k)\big) = \big(f^* \circ g^*\big)\alpha(v_1, \dots, v_k). \end{aligned}$$

3. $\mathrm{Id}_E^*(\alpha)(v_1, \dots, v_k) = \alpha\big(\mathrm{Id}_E(v_1), \dots, \mathrm{Id}_E(v_k)\big) = \alpha(v_1, \dots, v_k)$.
4. This follows from #2 and #3, since $f^* \circ (f^{-1})^* = (f^{-1} \circ f)^* = \mathrm{Id}_E^* = \mathrm{Id}_{\boldsymbol{\Omega}^*(E)}$.
5. Exercise □

B.16 Exercises (**Invariance of the Volume Form**)
On the symplectic vector space $(\mathbb{R}^{2n}, \omega)$ studied in Example B.8, show that linear symplectomorphisms (defined to be linear mappings $f : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ satisfying $f^*\omega = \omega$) leave the volume form $\bigwedge_{k=1}^{2n} \alpha_k$ invariant. ♢

B.2 Differential Forms on $\mathbb{R}^n$

We now introduce differential forms on manifolds, beginning with those on open subsets $U \subseteq \mathbb{R}^n$. This special case will describe the behavior of a general differential form on the domain of a chart.

A *differential form* ω on U is an exterior form that varies from point to point, in a manner that will be assumed to be smooth.

A general *k-form* ω is of the form

$$\omega \;=\; \sum_{1 \le i_1 < \dots < i_k \le n} \omega_{i_1 \dots i_k}\, dx_{i_1} \wedge \dots \wedge dx_{i_k} \;\in\; \Omega^k(U)\,, \tag{B.2.1}$$

where the $\omega_{i_1 \dots i_k}$ are functions from $C^\infty(U, \mathbb{R})$, and the dx_i are differential 1-forms associated with the coordinate functions $x_i : \mathbb{R}^n \to \mathbb{R}$ (namely $dx_i \in \Omega^1(\mathbb{R}^n)$). We here write this space, in contrast to the space $\mathbf{\Omega}^k(\mathbb{R}^n)$ of exterior k-forms, with a non-bold Ω.

The dx_i are defined by how they act on a vector field $v : U \to \mathbb{R}^n$, with

$$dx_i(v)(y) := v_i(y) \qquad (y \in U,\ i = 1, \dots, n).$$

Differential 1-forms thus map vector fields into functions; and for k vector fields $v^{(l)} : U \to \mathbb{R}$ and $\omega \in \Omega^k(U)$, one defines

$$\omega\left(v^{(1)}, \dots, v^{(k)}\right) := \sum_{1 \le i_1 < \dots < i_k \le n} \omega_{i_1 \dots i_k} \det \begin{pmatrix} dx_{i_1}(v^{(1)}) & \dots & dx_{i_k}(v^{(1)}) \\ \vdots & & \vdots \\ dx_{i_1}(v^{(k)}) & \dots & dx_{i_k}(v^{(k)}) \end{pmatrix}.$$

The result is again a real function on U. The rules of algebra from Section B.1 carry over from exterior forms to differential forms.

On $\Omega^*(U) := \bigoplus_{k=0}^n \Omega^k(U)$, we now consider the *differential operator* d, defined by

- $df := \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i$ for functions $f \in C^\infty(U, \mathbb{R}) = \Omega^0(U)$
- and $d\omega := \sum_{1 \le i_1 < \dots < i_k \le n} d\omega_{i_1 \dots i_k} \wedge dx_{i_1} \wedge \dots \wedge dx_{i_k}$ for the k-forms ω from (B.2.1).

Thus d converts a k-form into a $(k+1)$-form.

B.17 Definition *$d : \Omega^*(U) \to \Omega^*(U)$ is called the **exterior derivative**.*

B.18 Example (Exterior Derivative)

1. For $\omega := x_2\, dx_1 \in \Omega^1(\mathbb{R}^n)$,

$$d\omega = dx_2 \wedge dx_1 = -dx_1 \wedge dx_2\,.$$

2. For $\omega = \omega_1 dx_1 + \omega_2 dx_2 + \omega_3 dx_3 \in \Omega^1(\mathbb{R}^3)$,

$$\begin{aligned} d\omega &= (d\omega_1) \wedge dx_1 + (d\omega_2) \wedge dx_2 + (d\omega_3) \wedge dx_3 \\ &= \left(\tfrac{\partial \omega_2}{\partial x_1} - \tfrac{\partial \omega_1}{\partial x_2}\right) dx_1 \wedge dx_2 + \left(\tfrac{\partial \omega_3}{\partial x_2} - \tfrac{\partial \omega_2}{\partial x_3}\right) dx_2 \wedge dx_3 + \left(\tfrac{\partial \omega_1}{\partial x_3} - \tfrac{\partial \omega_3}{\partial x_1}\right) dx_3 \wedge dx_1\,. \end{aligned}$$

3. For $\omega = \omega_{12} dx_1 \wedge dx_2 + \omega_{23} dx_2 \wedge dx_3 + \omega_{31} dx_3 \wedge dx_1 \in \Omega^2(\mathbb{R}^3)$,

$$d\omega = \left(\tfrac{\partial \omega_{12}}{\partial x_3} + \tfrac{\partial \omega_{23}}{\partial x_1} + \tfrac{\partial \omega_{31}}{\partial x_2}\right) dx_1 \wedge dx_2 \wedge dx_3\,.$$

4. For $\omega \in \Omega^3(\mathbb{R}^3)$, one has $d\omega = 0$. ◊

B.19 Theorem (Exterior Derivative)
*d is an **antiderivation**, i.e., for $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^l(U)$, one has*

$$d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^k \alpha \wedge d\beta\,.$$

Proof: It suffices to prove this equation for monomials $\alpha := f\,\tilde{\alpha}$, $\beta := g\,\tilde{\beta}$ with $f, g \in C^\infty(\mathbb{R}^n, \mathbb{R})$, $\tilde{\alpha} := dx_{i_1} \wedge \ldots \wedge dx_{i_k}$ and $\tilde{\beta} := dx_{j_1} \wedge \ldots \wedge dx_{j_l}$, because d is linear. In this case,

$$\begin{aligned} d(\alpha \wedge \beta) &= d(f \cdot g)\tilde{\alpha} \wedge \tilde{\beta} = ((df)g + f(dg))\tilde{\alpha} \wedge \tilde{\beta} \\ &= (df)\tilde{\alpha} \wedge g\tilde{\beta} + (-1)^k f\tilde{\alpha} \wedge (dg)\tilde{\beta} = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta. \end{aligned}$$ □

B.20 Theorem $dd = 0$.

Proof:

1. For $f \in \Omega^0(U)$, one has $ddf = d\left(\sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i\right)$, so it equals

$$\sum_{i,l=1}^n \frac{\partial^2 f}{\partial x_l \partial x_i} dx_l \wedge dx_i = \sum_{1 \le r < s \le n} \left(\frac{\partial^2 f}{\partial x_r \partial x_s} - \frac{\partial^2 f}{\partial x_s \partial x_r}\right) dx_r \wedge dx_s = 0\,,$$

because the partial derivatives commute by the smoothness of f.

2. For $\omega = \sum \omega_{i_1 \ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k} \in \Omega^k(U)$, one finds by part 1 that

$$dd\omega = \sum (dd\omega_{i_1 \ldots i_k}) \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k} = 0\,.$$ □

Differential Operators, Coordinate Changes

We recall Example B.10, in which we converted vectors into 1-forms and $(n-1)$-forms respectively. We now want to do the same for vector fields and differential forms. So by means of the canonical scalar product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^b$, we associate with the vector field $v \in C^\infty(U, \mathbb{R}^n)$

1. the 1-form defined by $v^* \in \Omega^1(U)$, $v^*(w) := \langle v, w \rangle \quad \big(w \in C^\infty(U, \mathbb{R}^n)\big)$. In coordinates, $v^* = \sum_{i=1}^n v_i\, dx_i$.
2. an $(n-1)$-form $\omega_v \in \Omega^{n-1}(U)$; this assignment is defined by

$$\omega_v(w_1, \ldots, w_{n-1}) := \det(v, w_1, \ldots, w_{n-1}) \quad \big(w_i \in C^\infty(U, \mathbb{R}^n)\big)\,, \tag{B.2.2}$$

i.e., by application on $n-1$ vector fields. In coordinates, one has $\omega_v = dx_1 \wedge \ldots \wedge dx_n(v, \cdot, \ldots, \cdot)$, hence

$$\omega_v = \sum_{i=1}^n (-1)^{i-1} v_i dx_1 \wedge \ldots \wedge \widehat{dx_i} \wedge dx_{i+1} \wedge \ldots \wedge dx_n\,. \tag{B.2.3}$$

Here, $\widehat{dx_i}$ denotes omission of dx_i.

In the first case, one finds the rule

$$\boxed{\operatorname{grad}(f)^* = df} \tag{B.2.4}$$

for the *gradient*

$$\operatorname{grad}(f) \equiv \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

of a real function f, in the second case, one obtains

$$\boxed{\operatorname{div}(v)\, dx_1 \wedge \ldots \wedge dx_n = d\omega_v} \tag{B.2.5}$$

for the *divergence*

$$\operatorname{div}(v) \equiv \nabla \cdot v := \sum_{k=1}^{n} \frac{\partial v_k}{\partial x_k}$$

of a vector field v. Indeed, by (B.2.3),

$$d\omega_v = \sum_{i,k=1}^{n} (-1)^i \frac{\partial v_i}{\partial x_k} dx_k \wedge dx_1 \wedge \ldots \wedge \widehat{dx_i} \wedge dx_{i+1} \wedge \ldots \wedge dx_n,$$

and the terms in the sum vanish for $k \neq i$. Since $dx_1 \wedge \ldots \wedge dx_n$ is the canonical volume form on $\mathbb{R}^n$, we obtain a relation that is of practical use.

Specifically for $n = 3$ dimensions, the *curl*

$$\operatorname{rot} v \equiv \nabla \times v := \begin{pmatrix} \frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3} \\ \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1} \\ \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \end{pmatrix}$$

of the vector field v is connected to the exterior derivative by the relation

$$\boxed{\omega_{\operatorname{rot} v} = d(v^*)\,.} \tag{B.2.6}$$

- From (B.2.5), (B.2.6), and Theorem B.20, one obtains

  $$\operatorname{div}(\operatorname{rot} v)\, dx_1 \wedge dx_2 \wedge dx_3 = d\omega_{\operatorname{rot} v} = ddv^* = 0,$$

  hence (using that $dx_1 \wedge dx_2 \wedge dx_3 \neq 0$) the relation

  $$\boxed{\operatorname{div}\operatorname{rot}\ v{=}0\,.}$$

  This and similar relations are also valid if a different Riemannian metric (see page 172) is chosen, because they are derived from $dd = 0$.
- The relation

  $$\boxed{\operatorname{rot}\operatorname{grad} f = 0\,,}$$

which is true for arbitrary smooth functions f, also turns out to be a manifestation of the rule $dd = 0$: By (B.2.6) and (B.2.4), one has

$$\omega_{\text{rot}\,(\text{grad}\ f)} = d(\text{grad}\ f)^* = ddf = 0\,.$$

- As a last example for the usefulness of differential forms in vector analysis, we derive the identity

$$\boxed{\text{div}(v \times w) = \langle \text{rot}\, v, w\rangle - \langle v, \text{rot}\, w\rangle\ .}$$

We have already seen in Example B.10 that

$$\boxed{v^* \wedge w^* = \omega_{v\times w}} \tag{B.2.7}$$

holds: the corresponding rule for exterior forms carries over directly to differential forms in $\mathbb{R}^3$. So we conclude, using (B.2.5), (B.2.7), and (B.2.6), that

$$\begin{aligned}\text{div}(v \times w)\, dx_1 \wedge dx_2 \wedge dx_3 &= d\omega_{v\times w} = d(v^* \wedge w^*) = (dv^*) \wedge w^* - v^* \wedge dw^* \\ &= \omega_{\text{rot}\, v} \wedge w^* - v^* \wedge \omega_{\text{rot}\, w^*} \\ &= \big(\langle \text{rot}\, v, w\rangle - \langle v, \text{rot}\, w\rangle\big)\, dx_1 \wedge dx_2 \wedge dx_3\,,\end{aligned}$$

which was to be shown.

B.21 Example (Maxwell's Equations) The cartesian coordinates $x_1, \ldots, x_4$ on spacetime $\mathbb{R}^4$ denote the point $x := (x_1, x_2, x_3)$ in space and the time $t := x_4$. Let the *field strength* $F \in \Omega^2(\mathbb{R}^4)$ be given by

$$F := B_1\, dx_2 \wedge dx_3 + B_2\, dx_3 \wedge dx_1 + B_3\, dx_1 \wedge dx_2 + \sum_{i=1}^{3} E_i\, dx_i \wedge dx_4\,,$$

where $E := (E_1, E_2, E_3) \in C^\infty(\mathbb{R}^4, \mathbb{R}^3)$ and $B := (B_1, B_2, B_3) \in C^\infty(\mathbb{R}^4, \mathbb{R}^3)$ are the electric and magnetic field, respectively.

The *homogeneous Maxwell equation* $dF = 0$ is equivalent to

$$\text{div}_x(B) = 0 \quad , \quad \frac{\partial B}{\partial t} = -\text{rot}_x E\,.$$

From the Poincaré lemma B.45, we infer the existence of $A \in \Omega^1(\mathbb{R}^4)$ satisfying $F = dA$. This A is called a *gauge field.* ◊

Another aspect of differential forms is how they behave under mappings.

B.22 Definition *Let $U \subseteq \mathbb{R}^m$, $V \subseteq \mathbb{R}^n$ be open and $\varphi = \begin{pmatrix}\varphi_1\\ \vdots \\ \varphi_n\end{pmatrix} : U \to V$ smooth. The **pull-back** $\varphi^*\omega$ of a k-form*

$$\omega = \sum_{1\le i_1<\ldots<i_k\le n} \omega_{i_1\ldots i_k}\, dx_{i_1} \wedge \ldots \wedge dx_{i_k} \in \Omega^k(V)$$

on V is defined by

$$\varphi^*\omega = \sum_{1\le i_1<\ldots<i_k\le n} \omega_{i_1\ldots i_k} \circ \varphi \cdot d\varphi_{i_1} \wedge \ldots \wedge d\varphi_{i_k} \,.$$

So $\varphi^*\omega \in \Omega^k(U)$, in other words, the pull-back *is a k-form on $U \subseteq \mathbb{R}^m$.*

Just as the exterior derivative, the *pull-back* $\varphi^* : \Omega^*(V) \to \Omega^*(U)$ is a linear mapping.

B.23 Example (Area Form in Polar Coordinates)
The transition from cartesian to polar coordinates is achieved by

$$\varphi = \begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix} : \mathbb{R}^+ \times (0, 2\pi) \to \mathbb{R}^2 \quad , \quad \varphi_1(r,\psi) = r\cos(\psi)\,,\ \varphi_2(r,\psi) = r\sin(\psi)\,.$$

Let's pull back the 2-form $\omega = f\,dx_1 \wedge dx_2$. With $\tilde f := f \circ \varphi$, namely the function f written in polar coordinates, one obtains with $\varphi^*\omega = \tilde f\, d\varphi_1 \wedge d\varphi_2$ because of

$$d\varphi_1(r,\psi) = \cos(\psi)\,dr - r\sin(\psi)\,d\psi \quad , \quad d\varphi_2(r,\psi) = \sin(\psi)\,dr + r\cos(\psi)\,d\psi$$

that

$$\varphi^*\omega(r,\psi) = \tilde f(r,\psi)\; r\,(\cos^2\psi + \sin^2\psi)\,dr \wedge d\psi = \tilde f(r,\psi)\, r\, dr \wedge d\psi. \qquad \diamondsuit$$

B.24 Exercises (Pull-back of Exterior Forms) Let $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ be open, $\varphi \in C^\infty(U,V)$ and $\omega \in \Omega^k(V)$, $\psi \in \Omega^l(V)$. Show that

$$(\varphi^*\omega) \wedge (\varphi^*\psi) = \varphi^*(\omega \wedge \psi)\,. \qquad \diamondsuit$$

B.25 Theorem *$\varphi^* d = d\varphi^*$, so* pull-back *by $\varphi \in C^\infty(U,V)$ and exterior derivative commute.*

Proof: By linearity of d and φ^*, it suffices to apply both sides of the claimed identity to $f\,dx_{i_1} \wedge \ldots \wedge dx_{i_k}$. The left side yields

$$\begin{aligned}\varphi^* d(f\,dx_{i_1} \wedge \ldots \wedge dx_{i_k}) &= \varphi^* \textstyle\sum_{l=1}^n \mathrm{D}_l f\, dx_l \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k}\\ &= \underbrace{\textstyle\sum_{l=1}^n \mathrm{D}_l f \circ \varphi\, d\varphi_l}_{d(f\circ\varphi)} \wedge d\varphi_{i_1} \wedge \ldots \wedge d\varphi_{i_k}\,.\end{aligned}$$

Using the chain rule, it follows likewise for the right side that

$$d\varphi^*\big(f\,dx_{i_1}\wedge\ldots\wedge dx_{i_k}\big) = d\big(f\circ\varphi\,d\varphi_{i_1}\wedge\ldots\wedge d\varphi_{i_k}\big)$$
$$= d(f\circ\varphi)\wedge d\varphi_{i_1}\wedge\ldots\wedge d\varphi_{i_k}. \qquad \square$$

Specializing to diffeomorphisms φ, Theorem B.25 implies in particular that the exterior derivative is defined independent of the coordinate system used.

B.3 Integration of Differential Forms

We first integrate n-forms on $\mathbb{R}^n$, and then k-forms on k-dimensional submanifolds of $\mathbb{R}^n$.

B.26 Definition *Let $U\subseteq\mathbb{R}^n$ be open, and let $\omega\in\Omega^n(U)$ have compact support (i.e., for $\omega = f\,dx_1\wedge\ldots\wedge dx_n$, one has $f(x)=0$ outside a compact set $K\subset U$). Then the* **integral** *of ω is*

$$\int_U \omega := \int_U f(x)\,\mathrm{d}x_1\ldots\mathrm{d}x_n\,.$$

B.27 Theorem *Let $U, V\subseteq\mathbb{R}^n$ be open and $\varphi: V\to U$ a diffeomorphism with constant sign ε of $\det\big(\mathrm{D}\varphi(x)\big)$. Then*

$$\int_V \varphi^*\omega = \varepsilon\int_U \omega\,.$$

Proof: By definition of the pull-back, and using the symmetric group S_n, we calculate

$$\varphi^*\omega = f\circ\varphi\,d\varphi_1\wedge\ldots\wedge d\varphi_n = f\circ\varphi\sum_{i_1,\ldots,i_n=1}^{n}\frac{\partial\varphi_1}{\partial x_{i_1}}\cdot\ldots\cdot\frac{\partial\varphi_n}{\partial x_{i_n}}\,dx_{i_1}\wedge\ldots\wedge dx_{i_n}$$
$$= f\circ\varphi\sum_{\pi\in S_n}\frac{\partial\varphi_1}{\partial x_{\pi(1)}}\cdot\ldots\cdot\frac{\partial\varphi_n}{\partial x_{\pi(n)}}dx_{\pi(1)}\wedge\ldots\wedge dx_{\pi(n)}$$
$$= f\circ\varphi\left(\sum_{\pi\in S_n}\mathrm{sign}(\pi)\frac{\partial\varphi_1}{\partial x_{\pi(1)}}\cdot\ldots\cdot\frac{\partial\varphi_n}{\partial x_{\pi(n)}}\right)dx_1\wedge\ldots\wedge dx_n$$
$$= f\circ\varphi\,\det(\mathrm{D}\varphi)\,dx_1\wedge\ldots\wedge dx_n\,.$$

By the transformation theorem for integrals, the integration of this n-form on V yields
$\int_V f\circ\varphi\det(\mathrm{D}\varphi)\,\mathrm{d}x_1\ldots\mathrm{d}x_n = \varepsilon\int_V f\circ\varphi\,|\det\mathrm{D}\varphi|\,\mathrm{d}x = \varepsilon\int_U f\,\mathrm{d}x = \varepsilon\int_U\omega.$ □

We see in particular that the integral over the n-form ω does not depend on the choice of the (oriented) coordinate system.

If we consider $dx_1 \wedge \ldots \wedge dx_n$ as the *standard volume form* on $\mathbb{R}^n$, we can interpret $\int_U \omega$ also as an integral of the *function* f over U.

Next, if we want to integrate functions over k-dimensional manifolds $V \subseteq \mathbb{R}^n$ (so $k \leq n$), we first need to clarify the standard volume form of V. To this end, we use a parametrization of V: Let $U \subseteq \mathbb{R}^k$ be open, and

$$\varphi : U \to \mathbb{R}^n$$

an injective smooth mapping with image $\varphi(U) = V \subseteq \mathbb{R}^n$. Let

$$\operatorname{rank}\big(\mathrm{D}\varphi(x)\big) = k \qquad (x \in U),$$

i.e., the maximal rank. Then φ parametrizes the k-dimensional manifold V. We are now looking for a form $\omega^{(\varphi)} \in \Omega^k(U)$ for which

$$\int_{\varphi^{-1}(V')} \omega^{(\varphi)}$$

is the volume of V' for every $V' \subseteq V$ that is open in V.

Sensible requirements for a φ-dependent definition of $\omega^{(\varphi)}$ are the following:

- The square $V' := (0, 1)^k \times \{0\}^{n-k} \subset \mathbb{R}^n$ should have area 1.
- Under a rotation $O \in \mathrm{SO}(n)$, the area of V' should not change, and not under translations either.
- The area of V should not change under an orientation-preserving change of parameters.

The **volume form**

$$\omega^{(\varphi)} := \sqrt{\det(g)}\, dx_1 \wedge \ldots \wedge dx_k \tag{B.3.1}$$

of the parametrized surface V meets these requirements, if the symmetric $k \times k$ matrix g is defined by

$$\boxed{g := (\mathrm{D}\varphi)^\top \mathrm{D}\varphi\,.}$$

For a regular parametrization ($\operatorname{rank}\big(\mathrm{D}\varphi(x)\big) = k \quad (x \in U)$), the matrix g is positive definite, $g(x) > 0$, because

$$\langle v, g(x)v\rangle = \langle \mathrm{D}\varphi(x)\, v,\ \mathrm{D}\varphi(x)\, v\rangle = \|\mathrm{D}\varphi(x)\, v\|_2^2 > 0 \qquad \big(v \in \mathbb{R}^k \setminus \{0\}\big).$$

g is therefore a metric tensor in the sense of Definition A.51. In the literature, $\det(g)$ is often denoted as $|g|$.

As an example, let us show the invariance of g under rotation: If $\psi : V \to \mathbb{R}^n$ is given by $\psi = O \circ \varphi$ with $O \in \mathrm{SO}(n)$, then

$$(\mathrm{D}\psi)^\top \mathrm{D}\psi = (O\mathrm{D}\varphi)^\top (O\mathrm{D}\varphi) = (\mathrm{D}\varphi)^\top O^\top O\mathrm{D}\varphi = (\mathrm{D}\varphi)^\top \mathrm{D}\varphi = g\,.$$

We can integrate a function $f : V \to \mathbb{R}$ by forming the integral

$$\boxed{\int_U f \circ \varphi \cdot \omega^{(\varphi)}\,.}$$

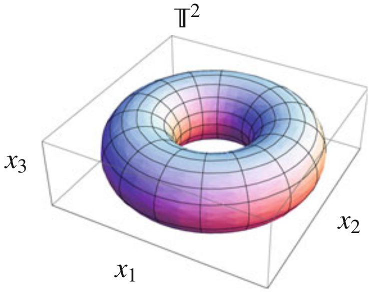
This expression is invariant under change in parametrization. The special case $f = \mathbb{1}_V$ yields again the k-dimensional volume of V.

B.28 Example (Volume) We calculate the surface area of a two dimensional torus $\mathbb{T}^2 \subset \mathbb{R}^3$. It is parametrized, for radii $r_1 > r_2 > 0$, by $\mathbb{T}^2 := \varphi(U)$ with

$$\varphi : U \to \mathbb{R}^3\,,\ U := [0, 2\pi) \times [0, 2\pi)$$

and

$$\varphi(\psi_1, \psi_2) := \begin{pmatrix} (r_1+r_2\cos\psi_2)\cos\psi_1 \\ (r_1+r_2\cos\psi_2)\sin\psi_1 \\ r_2\sin\psi_2 \end{pmatrix}.$$



We ignore the fact that $U \subset \mathbb{R}^2$ is not open, because the boundary of U is a set of Lebesgue measure zero.

The coefficients of the Riemannian metric $g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$ with $g_{21} = g_{12}$ are

$$\begin{aligned}
g_{11}(\psi_1, \psi_2) &= \sum_{i=1}^{3} \left(\frac{\partial \varphi_i}{\partial \psi_1}\right)^2 = (r_1 + r_2\cos\psi_2)^2(\sin^2\psi_1 + \cos^2\psi_1) \\
&= (r_1 + r_2\cos\psi_2)^2\,, \\
g_{12}(\psi_1, \psi_2) &= \sum_{i=1}^{3} \frac{\partial \varphi_i}{\partial \psi_1}\frac{\partial \varphi_i}{\partial \psi_2} = r_2\sin\psi_2(r_1 + r_2\cos\psi_2)\sin\psi_1\cos\psi_1 \\
&\qquad\qquad -r_2\sin\psi_2(r_1 + r_2\cos\psi_2)\cos\psi_1\sin\psi_1 = 0\,, \\
g_{22}(\psi_1, \psi_2) &= \sum_{i=1}^{3} \left(\frac{\partial \varphi_i}{\partial \psi_2}\right)^2 = r_2^2[\sin^2\psi_2(\cos^2\psi_1 + \sin^2\psi_1) + \cos^2\psi_1] = r_2^2,
\end{aligned}$$

hence $\sqrt{|g|} = \sqrt{\det g} = r_2(r_1 + r_2\cos\psi_2) > 0$; thus the surface of the torus is

$$\int_U \omega^{(\varphi)} = \int_0^{2\pi}\int_0^{2\pi} \sqrt{\det g}\,\mathrm{d}\psi_1\,\mathrm{d}\psi_2 = (2\pi)^2 r_1 r_2\,. \qquad \diamondsuit$$

Let us now have a closer look at the special case (as in the example) of a *hypersurface* V in $\mathbb{R}^n$. So $V = \varphi(U) \subset \mathbb{R}^n$ has a parametrization

$$\varphi : U \to \mathbb{R}^n \quad \text{for } U \subseteq \mathbb{R}^{n-1} \text{ open.}$$

There exists a continuous *normal vector field*

$$N : V \to \mathbb{R}^n \quad , \quad \|N(y)\| = 1 \qquad (y \in V)$$

on V (unique up to sign), which is orthogonal on V, so

$$\langle N \circ \varphi(x), \mathrm{D}\varphi(x)\, w \rangle = 0 \qquad \big(x \in U,\ w \in \mathbb{R}^{n-1}\big).$$

Then, for the $n \times n$ matrix $M(x) := \big(N \circ \varphi(x), \mathrm{D}\varphi(x)\big)$, one has

$$M^\top(x)\, M(x) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \mathrm{g(x)} & \\ 0 & & & \end{pmatrix} \qquad (x \in U),$$

and therefore $\det g(x) = \det\big(M^\top(x) M(x)\big) = \big(\det(M(x))\big)^2$, or, with the appropriate orientation of the normal vector field,

$$\sqrt{\det g} = \det M. \tag{B.3.2}$$

From this one obtains, for the $(n-1)$-form ω_N on V that is dual to the normal vector field N and defined in (B.2.2), that ω_N is equal to the volume form of the hypersurface V:

B.29 Theorem *$\varphi^*\omega_N = \omega^{(\varphi)}$ for $\omega^{(\varphi)}$ from (B.3.1).*

Proof: Again denoting by $\widehat{d\varphi_k}$ the omission of $d\varphi_k$, we have

$$\varphi^*\omega_N = \sum_{k=1}^{n} (-1)^{k-1} N_k \circ \varphi\, d\varphi_1 \wedge \ldots \wedge \widehat{d\varphi_k} \wedge \ldots d\varphi_n = \det M\, dx_1 \wedge \ldots \wedge dx_{n-1}.$$

The claim follows from (B.3.2). Strictly speaking, by definition of the *pull-back*, the $(n-1)$-form ω_N should be defined on an open neighborhood of V in $\mathbb{R}^n$. This condition can be met by smooth extension of the normal vector field N, and $\varphi^*\omega_N$ does then not depend on the choice of the extension. □

Another important situation occurs when in the space around the k-dimensional submanifold $V \subseteq \mathbb{R}^n$, a k-form $\omega \in \Omega^k(\mathbb{R}^n)$ is defined. Its integral over the submanifold $V = \varphi(U)$ (parametrized by an open subset $U \subset \mathbb{R}^k$ and a chart $\varphi : U \to \mathbb{R}^n$) is defined by

$$\int_V \omega := \int_U \varphi^*\omega. \tag{B.3.3}$$

According to Theorem B.27, this integral is independent, except for a sign, of the chosen parametrization.

B.4 Differential Forms on Manifolds

If in Definition B.12 of the Grassmann algebra $\mathbf{\Omega}^*(E)$, we choose the $\mathbb{R}$-vector space E to be the tangent space $T_m M$ of the manifold M at $m \in M$, then the elements $\omega(m) \in \mathbf{\Omega}^*(T_m M)$ are exterior forms at m.
Smoothness of $m \mapsto \omega(m)$ can be defined by means of the atlas of M:

B.30 Definition *On the manifold M with atlas $\{(U_i, \varphi_i) \mid i \in I\}$, a **differential form** ω is a family of differential forms in the images of the charts,*

$$\omega_i \in \Omega^*(V_i) \quad , \quad V_i := \varphi_i(U_i) \subseteq \mathbb{R}^n \qquad (i \in I)$$

that are compatible in the following sense: For $i, j \in I$ and the domain $V_{i,j} := \varphi_i(U_i \cap U_j) \subset V_i$ of the coordinate change map $\psi_{i,j} := \varphi_j \circ \varphi_i^{-1}\restriction_{V_{i,j}} : V_{i,j} \to V_{j,i}$, one has

$$\psi_{i,j}^*\big(\omega_j\restriction_{V_{j,i}}\big) = \omega_i\restriction_{V_{i,j}} \qquad (i, j \in I).$$

With addition defined in each chart, one obtains the $\mathbb{R}$-vector space

$$\Omega^*(M) = \bigoplus_{k=0}^{n} \Omega^k(M) \tag{B.4.1}$$

of the differential forms on the manifolds M, and the $\wedge$ product is also defined by charts.
The pull-back of differential forms given in Definition B.22 also carries over from open subsets of $\mathbb{R}^n$ to manifolds:
So let $f : M \to N$ be a smooth mapping between the manifolds M, N, and let $\omega \in \Omega^k(N)$ be a k-form. For $m \in M$, tangent vectors $u_1, \ldots, u_k \in T_m M$, and their images $v_i := T_m f(u_i) \in T_{f(m)} N$, the *pull-back* $f^*\omega \in \Omega^k(M)$ is defined by

$$f^*\omega(m)(u_1, \ldots, u_k) := \omega\big(f(m)\big)(v_1, \ldots, v_k).$$

So we can use Definition B.22 in the local representations (A.2.1) of f. Applying the commutability of exterior derivative and diffeomorphism (Theorem B.25) to the coordinate change maps, we see that the *exterior derivative*

$$d : \Omega^k(M) \to \Omega^{k+1}(M) \qquad (k \in \mathbb{N}_0)$$

of differential forms on manifolds M is well-defined, independent of coordinate charts.

B.5 Inner Derivative and Lie Derivative

The *Lie derivative* of a differentiable function $f : M \to \mathbb{R}$ on a manifold M in the direction of a vector field $X : M \to TM$ is the real function on M defined by

$$L_X f := df(X)\,. \tag{B.5.1}$$

Generalizing, we will also define the Lie derivative of a k-form with respect to X. To this end, a notation is convenient in which the pairing of differential forms and vector fields is written without reversing the order, in contradistinction to the right side of (B.5.1).

B.31 Definition *Let X be a vector field on a manifold M and ω a $(k+1)$-form on M. Then the k-form $\mathbf{i}_X\omega \in \Omega^k(M)$ defined by*

$$\mathbf{i}_X\omega(X_1, \dots, X_k) := \omega(X, X_1, \dots, X_k)$$

is called the ***inner product*** *of X and ω.*

Of course one can also write $\omega(X, \cdot, \dots, \cdot)$ instead of $\mathbf{i}_X\omega$.

B.32 Exercises (**$\wedge$-antiderivation**) Show that for vector fields X on M, the mapping $\mathbf{i}_X$ is a **$\wedge$-antiderivation**, i.e., it is linear and

$$\mathbf{i}_X(\varphi \wedge \omega) = (\mathbf{i}_X\varphi) \wedge \omega + (-1)^k \varphi \wedge \mathbf{i}_X(\omega) \quad \big(\varphi \in \Omega^k(M),\ \omega \in \Omega^l(M)\big). \tag{B.5.2}$$

B.33 Definition *Let X be a vector field on a manifold M. Then for $\omega \in \Omega^k(M)$, the* ***Lie derivative*** *of ω with respect to X is defined to be the k-form*

$$L_X\omega := (\mathbf{i}_X d + d\,\mathbf{i}_X)\,\omega\,.$$

Indeed, $L_X\omega \in \Omega^k(M)$, because the exterior derivative raises the degree of the form by 1, and taking the inner product lowers it by 1.

Definition B.33 generalizes Definition (B.5.1) of the Lie derivative of a function $f : M \to \mathbb{R}$ in direction X, because

$$L_X f = \mathbf{i}_X df + d\,\mathbf{i}_X f = \mathbf{i}_X df = df(X)\,,$$

as the inner product of a vector field with a function is 0 by definition (there are no differential forms of degree -1).

The Lie derivative $L_X\omega$ has a geometric interpretation. Namely, it describes the change of the differential form ω in the direction of the flow generated by the vector field X.

B.34 Theorem *Let the vector field X on a manifold M generate a flow $\Phi_t : M \to M$ $(t \in \mathbb{R})$. Then for all $\omega \in \Omega^*(M)$, one has*

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi_t^*\omega\right) = \Phi_t^* L_X\omega \quad (t \in \mathbb{R}).}$$

Proof:

- We only need to prove the relation for $t = 0$, because

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_t^*\omega) = \frac{\mathrm{d}}{\mathrm{d}s}\Phi_{t+s}^*\omega\Big|_{s=0} = \frac{\mathrm{d}}{\mathrm{d}s}\Phi_t^*\Phi_s^*\omega\Big|_{s=0} = \Phi_t^*\left(\frac{\mathrm{d}}{\mathrm{d}s}\Phi_s^*\omega\right)\Big|_{s=0}.$$

- We begin with the Lie derivative of functions $f : M \to \mathbb{R}$. In local coordinates $y = (y_1, \ldots, y_n)$, one finds $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t^* f(y)\big|_{t=0}$ to be equal to

$$\lim_{t\to 0}\frac{f\big(\Phi_t(y)\big) - f(y)}{t} = \sum_{i=1}^{n}\frac{\partial f}{\partial y_i}(y)\cdot X_i(y) = df(X)(y) = L_X f(y).$$

- If we choose for ω the special 1-forms dy_i, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_t^* dy_i)\big|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}d(\Phi_t^* y_i)\big|_{t=0} = d\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_t^* y_i)\big|_{t=0} = dX_i\,,$$

  and on the other hand, we get from $L_X d = \mathbf{i}_X dd + d\,\mathbf{i}_X d = d\,\mathbf{i}_X d = dL_X$ that

$$L_X dy_i = dL_X y_i = d\,\mathbf{i}_X dy_i = dX_i\,.$$

- The theorem now follows from $\Phi_t^*(\varphi\wedge\omega) = (\Phi_t^*\varphi)\wedge(\Phi_t^*\omega)$ and the following lemma, since every k-form ω can be written in local coordinates $y = (y_1, \ldots, y_n)$ as

$$\omega = \sum_{1\le l_1<\ldots<l_k\le n} f_{l_1\ldots l_k}\,dy_{l_1}\wedge\ldots\wedge dy_{l_k}\,.$$ □

The Lie derivative is a derivation on the algebra $\Omega^*(M)$ of differential forms:

B.35 Lemma *For φ, $\omega \in \Omega^*(M)$, one has $L_X(\varphi\wedge\omega) = (L_X\varphi)\wedge\omega + \varphi\wedge L_X\omega$.*

Proof: For k-forms φ, we have

$$d(\varphi\wedge\omega) = (d\varphi)\wedge\omega + (-1)^k\varphi\wedge d\omega \quad\text{and}\quad \mathbf{i}_X(\varphi\wedge\omega) = (\mathbf{i}_X\varphi)\wedge\omega + (-1)^k\varphi\wedge\mathbf{i}_X\omega$$

according to Theorem B.19 and (B.5.2). Combining both antiderivations yields the Leibniz rule

$$\begin{aligned}L_X(\varphi\wedge\omega) &= \mathbf{i}_X d(\varphi\wedge\omega) + d\,\mathbf{i}_X(\varphi\wedge\omega)\\ &= \mathbf{i}_X\big((d\varphi)\wedge\omega + (-1)^k\varphi\wedge d\omega\big) + d\big((\mathbf{i}_X\varphi)\wedge\omega + (-1)^k\varphi\wedge\mathbf{i}_X\omega\big)\\ &= (L_X\varphi)\wedge\omega + \varphi\wedge L_X\omega\,.\end{aligned}$$ □

B.36 Remark (Coordinate Vector Fields) With reference to a coordinate system $\varphi_1, \ldots, \varphi_n : U \to \mathbb{R}$, we have the *coordinate vector fields*

$$\frac{\partial}{\partial \varphi_i} \equiv \partial_{\varphi_i} : U \to \mathbb{R}^n \qquad (i = 1, \ldots, n),$$

which are determined by

$$L_{\partial_{\varphi_i}} \varphi_k := \delta_{ik} \qquad (i, k = 1, \ldots, n)\,, \tag{B.5.3}$$

and in whose direction only the i^{th} coordinate gets varied. They satisfy

$$L_{\partial_{\varphi_i}} g(\varphi_1, \ldots, \varphi_n) = \frac{\partial}{\partial \varphi_i} g(\varphi_1, \ldots, \varphi_n)\,,$$

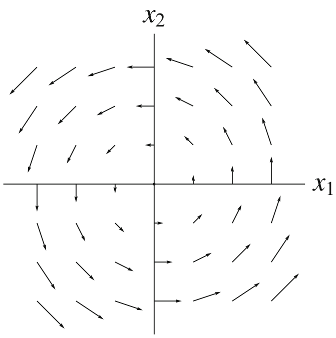
so the Lie derivative of the function $g \in C^1(U, \mathbb{R})$ in the direction of the i^{th} vector field is therefore equal to the i^{th} partial derivative; hence the notation $\frac{\partial}{\partial \varphi_i}$.

Note for $n > 1$ that in view of (B.5.3), it does not suffice to know the i^{th} coordinate φ_i in order to calculate $\frac{\partial}{\partial \varphi_i}$. ♢

B.37 Example (Coordinate Vector Fields for Polar Coordinates) On the open subset $U := \mathbb{R}^2 \backslash \big((-\infty, 0] \times \{0\}\big)$ of $\mathbb{R}^2$ (plane with a slit), we define polar coordinates

$$\varphi(x) := \arctan\left(\frac{x_2}{x_1}\right) \quad , \quad r(x) := \sqrt{x_1^2 + x_2^2}$$

(where the angle $\varphi(x)$ is extended continuously for $x_2 \leq 0$, and thus $\varphi(x) \in (-\pi, \pi)$).



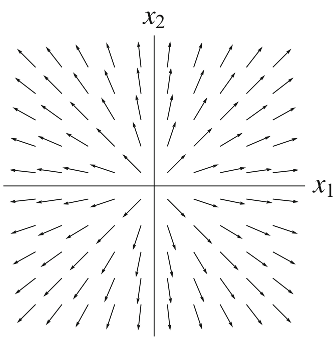
In cartesian coordinates, the vector field $\frac{\partial}{\partial \varphi}$ is of the form $\frac{\partial}{\partial \varphi}(x) = \binom{-x_2}{x_1}$ (see figure on the right), because

$$0 = L_{\frac{\partial}{\partial \varphi}} r = (x_1/r, x_2/r) \binom{-x_2}{x_1} \quad \text{and}$$

$$1 = L_{\frac{\partial}{\partial \varphi}} \varphi = \left(\frac{-x_2/x_1^2}{1+(x_2/x_1)^2}, \frac{1/x_1}{1+(x_2/x_1)^2}\right) \binom{-x_2}{x_1}.$$

Analogously, the radial vector field is

$$\frac{\partial}{\partial r}(x) = \begin{pmatrix} x_1/r(x) \\ x_2/r(x) \end{pmatrix} \,.$$



(See the figure on the right.) ♢

B.6 Stokes' Theorem

Let us suppose that in the space surrounding the k-dimensional submanifold $V \subseteq \mathbb{R}^n$, a k-form $\omega \in \Omega^k(\mathbb{R}^n)$ is defined. Its integral over the submanifold $V = \varphi(U)$ (parametrized by some open subset $U \subseteq \mathbb{R}^k$ and $\varphi : U \to \mathbb{R}^n$) was defined in (B.3.3) as

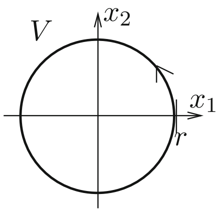
$$\int_V \omega := \int_U \varphi^* \omega.$$

By Theorem B.27, this integral is independent of the parametrization, except for a sign.

B.38 Example (Integral along a Path)
We integrate the 1-form

$$\omega \in \Omega^1(\mathbb{R}^2) \quad , \quad \omega(x) := x_1 dx_2$$

along the following path defined on $U := [0, 2\pi)$:



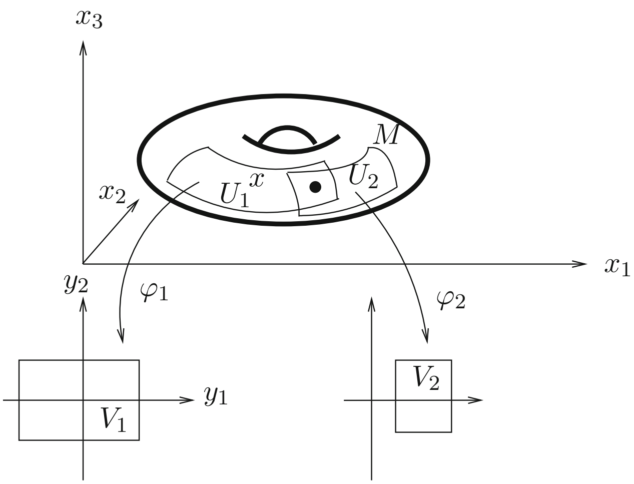
$$\varphi : U \to \mathbb{R}^2 \quad , \quad \varphi(\psi) := \begin{pmatrix} r\cos\psi \\ r\sin\psi \end{pmatrix}.$$

Its image $V := \varphi(U)$ is the circle of radius $r > 0$ about the origin.

$$\int_V \omega = \int_U \varphi^* \omega = \int_U r\cos\psi \, d(r\sin\psi) = r^2 \int_U \cos^2\psi \, d\psi = \pi r^2. \tag{B.6.1}$$

It is notable in this example that the integral of $d\omega = dx_1 \wedge dx_2$, namely the canonically oriented area element of $\mathbb{R}^2$, over the disc of radius r bounded by V equals the integral (B.6.1) of ω over the circle.

This is not an accident, but is a manifestation of a general theorem, called Stokes' theorem, which relates integrals over manifolds with integrals over their boundary.



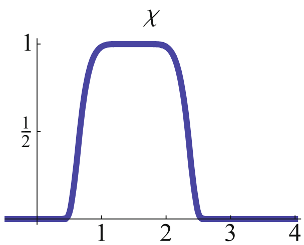
The fundamental theorem of calculus possesses the following generalization:

B.39 Theorem (Stokes) *Let M be an (oriented) k-dimensional manifold with boundary, and ω a $(k-1)$-form with compact support*[9] *on M. Then*

$$\boxed{\int_M d\omega = \int_{\partial M} \omega.}$$

Proof:

- Here we only give the proof for the case of a submanifold with boundary of $M \subseteq \mathbb{R}^n$ that has the full dimension $k = n$. The proof for arbitrary, not necessarily imbedded, manifolds with boundary can be found, e.g., in AGRICOLA and FRIEDRICH [AF], Chapter 3.6.
- By the implicit function theorem, we find for every point $x \in \partial M$ a neighborhood $U_i \subseteq M$ of x and a coordinate chart $\varphi_i : U_i \to V_i \subseteq \mathbb{R}^n_+$ such that $M \cap U_i = \varphi_i^{-1}(\mathbb{R}^{n-1} \times \{0\})$. As ω is non-zero only on some compact set, finitely many coordinate charts are sufficient.



- We can now employ a trick to use the seemingly restrictive hypothesis $\text{supp}(\omega) \subseteq U_i$. Namely, there is a smooth *partition of unity* (see Definition A.13), a family of functions $\chi_i \in C^\infty(\mathbb{R}^n, \mathbb{R})$ with $\chi_i \geq 0$, $\text{supp}(\chi_i) \subset U_i$ and $\sum_{i \in I} \chi_i = 1$, see the figure. If we now set $\omega_i := \chi_i \omega$, then $\sum_{i \in I} \omega_i = \omega$ and $\text{supp}(\omega_i) \subset U_i$.

[9] Alternatively one can stipulate decay conditions for ω and $d\omega$ to guarantee the finiteness of both sides.

- By the above definition, it suffices to consider the integration of an $(n-1)$-form ω whose support is contained in the half space $\mathbb{R}^n_+$. Then ω can be written (we again indicate the omission of a 1-form by a hat, as in Example A.29.2) in the form

$$\omega = \sum_{k=1}^{n} (-1)^{k-1} f_k \, dx_1 \wedge \ldots \wedge dx_{k-1} \wedge \widehat{dx_k} \wedge dx_{k+1} \wedge \ldots \wedge dx_n \, ,$$

where the f_k are functions with compact support in $\mathbb{R}^n_+$.
- Now $d\omega = \left(\sum_{k=1}^n \frac{\partial f_k}{\partial x_k}\right) dx_1 \wedge \ldots \wedge dx_n$, hence

$$\begin{aligned}\int_{\mathbb{R}^n_+} d\omega &= \sum_{k=1}^{n} \int_{\mathbb{R}^n_+} \frac{\partial f_k}{\partial x_k} \, dx_1 \wedge \ldots \wedge dx_n \\ &= \sum_{k=1}^{n-1} (-1)^{k-1} \int_{\mathbb{R}^{n-1}_+} \left(\int_{\mathbb{R}} \frac{\partial f_k}{\partial x_k} \, dx_k \right) dx_1 \wedge \ldots \wedge dx_{k-1} \wedge \widehat{dx_k} \wedge dx_{k+1} \wedge \ldots \wedge dx_n \\ &\quad + (-1)^{n-1} \int_{\mathbb{R}^{n-1}} \left(\int_0^\infty \frac{\partial f_n}{\partial x_n} \, dx_n \right) dx_1 \wedge \ldots \wedge dx_{n-1} \, .\end{aligned}$$

The inner integral vanishes for $k = 1, \ldots, n-1$ by the fundamental theorem of calculus. For the last term in the sum, however, integration by parts yields

$$\int_0^\infty \frac{\partial f_n}{\partial x_n}(x_1, \ldots, x_n) \, dx_n = -f_n(x_1, \ldots, x_{n-1}, 0) \, ,$$

and thus altogether

$$\int_{\mathbb{R}^n_+} d\omega = (-1)^n \int_{\mathbb{R}^{n-1}} f_n(x_1, \ldots, x_{n-1}, 0) \, dx_1 \wedge \ldots \wedge dx_{n-1} \, .$$

- On the other hand, $\omega\restriction_{\mathbb{R}^{n-1}\times\{0\}} = (-1)^{n-1} f_n \, dx_1 \wedge \ldots \wedge dx_{n-1}$, because x_n vanishes on this subspace, and thus $dx_n\restriction_{\mathbb{R}^{n-1}\times\{0\}} = 0$ as well. With the right choice of orientation, one therefore obtains $\int_{\mathbb{R}^n_+} d\omega = \int_{\mathbb{R}^{n-1}\times\{0\}} \omega$, which is Stokes' Theorem. □

B.40 Example (Stokes' Theorem)

1. Let $M \subset \mathbb{R}^n$ be the image of a (regular, injective) curve

$$c : [0, 1] \to \mathbb{R}^n$$

and $F : M \to \mathbb{R}$ a smooth function.

Then ∂M consists of the points $c(0)$ and $c(1)$. However, as start and end points, they have different orientation, and this gives rise to the formula

$$\int_0^1 \underbrace{\frac{\mathrm{d}}{\mathrm{d}t} F \circ c(t)\,\mathrm{d}t}_{dF\circ c} = F \circ c(1) - F \circ c(0)\,. \tag{B.6.2}$$

If $F = \tilde{F}\restriction_M$ for some smooth function $\tilde{F} : \mathbb{R}^n \to \mathbb{R}$, then the integral in (B.6.2) equals $\tilde{F}(c(1)) - \tilde{F}(c(0))$, independent of the choice of the path c connecting the prescribed start and end points.

2. Generalizing Example B.38, we consider the symplectic 2-form

$$\omega = -d\theta = \sum_{i=1}^{n} dq_i \wedge dp_i$$

on the phase space $P := \mathbb{R}^n_p \times \mathbb{R}^n_q$ with the 1-form $\theta := -\sum_{i=1}^{n} p_i dq_i$.
Let $c : S^1 \to P$ be a loop whose image bounds the image M of a disc, i.e.,

$$c(S^1) = \partial M\,.$$

Then the integral $\int_{\partial M} \theta = -\int_M \omega$ is independent of the choice of M. This quantity plays an important role when calculating action variables in integrable Hamiltonian systems.

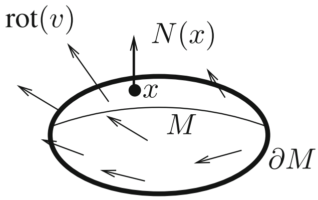
3. We stay with the image M of a disc $\tilde{M} \subset \mathbb{R}^2$. This time however, the disc is imbedded in $\mathbb{R}^3$, namely with an imbedding $\imath : \tilde{M} \to M \subset \mathbb{R}^3$.
Moreover, let $v : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. Then v^* is a 1-form, and dv^* a 2-form (thus integrable over M), and by the formula from (B.2.6),

$$\omega_{\mathrm{rot}(v)} = dv^*\,.$$

Hence the integral of dv^* over M equals the integral of the scalar product of $\mathrm{rot}(v)$ and the surface normal N (with respect to the surface element $\sqrt{|g|}dx_1 \wedge dx_2$). Therefore, by Theorem B.29, the **Kelvin-Stokes Theorem** follows:

$$\int_{\tilde{M}} \langle \mathrm{rot}(v), N\rangle \circ \imath(x)\, \sqrt{|g(x)|}\, dx_1 \wedge dx_2 = \int_M \langle \mathrm{rot}(v), N\rangle\, \omega_N$$
$$= \int_M \omega_{\mathrm{rot}(v)} = \int_M dv^* = \int_{\partial M} v^* = \int_{t_0}^{t_1} \left\langle v\big(c(t)\big), \frac{\mathrm{d}c}{\mathrm{d}t}(t)\right\rangle \mathrm{d}t$$

for a parametrization $c : [t_0, t_1] \to \partial M$ of the (oriented) boundary of the disc. The integral along the path is also called *circulation*.



For example, v might be the velocity field of a liquid. Then the above formula shows, if the flow is irrotational, that the component of the velocity that is tangential to the loop ∂M averages to 0.

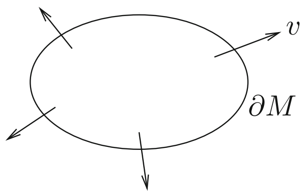
4. The divergence $\operatorname{div}(v)$ of a vector field v on $\mathbb{R}^n$ is related to the exterior derivative by (B.2.5) via $\omega_v = \sum_{i=1}^n (-1)^{i-1} v_i \, dx_1 \wedge \ldots \wedge \widehat{dx_i} \wedge dx_{i+1} \wedge \ldots \wedge dx_n$ and

$$d\,\omega_v = \operatorname{div}(v)\, dx_1 \wedge \ldots \wedge dx_n \,.$$

If $M \subset \mathbb{R}^n$ is an n-dimensional manifold with boundary (say a ball), then Stokes' theorem says

$$\int_M d\omega_v = \int_{\partial M} \omega_v \,.$$

For the boundary points $x \in \partial M$, let $N(x)$ denote the normal vector. $v(x)$ can be expressed uniquely in the form



$$v(x) = \langle v(x), N(x) \rangle \, N(x) + w(x) \,, \tag{B.6.3}$$

where $w(x)$ is tangential to ∂M at x, hence $\int_{\partial M} \omega_w = 0$, and with (B.6.3), $\int_{\partial M} \omega_v = \int_{\partial M} \langle v, N \rangle \, \omega_N$. One obtains **Gauss' Theorem**

$$\int_M \operatorname{div}(v)\, dx_1 \ldots dx_n = \int_M d\omega_v = \int_{\partial M} \omega_v = \int_{\partial M} \langle v, N \rangle \, \omega_N \,.$$

If the vector field is divergence free (as is for instance the velocity field of an incompressible fluid), then as much will exit from M through the boundary ∂M as will enter. ◊

B.7 The Poincaré Lemma

B.41 Definition *A k-form $\alpha \in \Omega^k(M)$ is called* ***closed****, if $d\alpha = 0$.*
It is called ***exact****, if there exists a differential form $\beta \in \Omega^{k-1}(M)$ with $\alpha = d\beta$.*

Because $dd = 0$ (Theorem B.20), every exact form is closed. However, not every closed from is exact:

B.42 Example (Aharonov-Bohm Effect) On $U := \mathbb{R}^2 \backslash \{0\}$, the 1-form $\omega := \frac{x_1 dx_2 - x_2 dx_1}{x_1^2 + x_2^2}$ is smooth, where x_1, x_2 are the cartesian coordinates on $\mathbb{R}^2$.

- This 1-form ω is closed because

$$d\omega = \left(\mathrm{D}_1 \left(\frac{x_1}{x_1^2 + x_2^2} \right) + \mathrm{D}_2 \left(\frac{x_2}{x_1^2 + x_2^2} \right) \right) dx_1 \wedge dx_2 = 0 \,.$$

- But ω is not exact. For if there were a 0-form $\varphi : U \to \mathbb{R}$ with $d\varphi = \omega$, then Stokes' theorem would imply for the integral over the circle $S^1 \subset U$:

$$\int_{S^1} d\varphi = \int_{\partial S^1} \varphi = 0 \,,$$

  because as the boundary of the disc, the circle S^1 itself has no boundary. On the other hand, we calculate:

$$\begin{aligned} \int_{S^1} d\varphi &= \int_{S^1} \omega = \int_{S^1} (x_1\, dx_2 - x_2\, dx_1) \\ &= \int_0^{2\pi} (\cos\varphi\, d(\sin\varphi) - \sin\varphi\, d(\cos\varphi)) = \int_0^{2\pi} (\cos^2\varphi + \sin^2\varphi)\, d\varphi = 2\pi \,. \end{aligned}$$

- The 1-form ω can be viewed as the gauge potential of a shielded (infinitely thin) solenoid along the x_3-axis. The exterior derivative $B := d\omega$ of the gauge field is the magnetic field, and it vanishes in the exterior domain.
- Even though no magnetic forces act on particles in the exterior domain, the quantum mechanical nature of electrons allows to verify experimentally that the gauge field ω itself is not zero. In an appropriate layout of the experiment, interference patterns are obtained (*Aharonov-Bohm effect*). ◊

However, if the domain U is star-shaped (or more generally: contractible), then every closed k-form on U is exact ($k \geq 1$).

B.43 Definition *An open subset $U \subseteq \mathbb{R}^n$ is called* ***star-shaped*** *if there exists an $x \in U$ such that for all $y \in U$, the segment between x and y lies in U.*

B.44 Example Open convex subsets $\emptyset \neq U \subseteq \mathbb{R}^n$ are star-shaped. ◊

B.45 Theorem (Poincaré Lemma) *If $U \subseteq \mathbb{R}^n$ is a star-shaped domain and $\omega \in \Omega^k(U)$, $k \geq 1$ is closed ($d\omega = 0$), then ω is exact ($\omega = d\varphi$).*

Proof:
While φ is not determined uniquely, the proof (following Abraham and Marsden [AM], *Theorem 2.4.17*) will contain a formula for one such $(k-1)$-form φ.

- We assume without loss of generality that U is star-shaped with respect to the origin.
- For $t \in (0, 1]$, the scaling maps

$$F_t : U \to U \quad , \quad y \mapsto ty \qquad (t \in [0, 1])$$

are diffeomorphisms onto their image $F_t(U)$. The mappings $t \mapsto F_t(u)$ for $u \in U$ can be viewed as solutions to the initial value problem

$$\frac{\mathrm{d}y}{\mathrm{d}t} = X_t(y) \quad , \quad y(1) = u$$

with the time-dependent vector field $X : (0, 1] \times \mathbb{R}^n \to \mathbb{R}^n$, $X_t(y) := y/t$. Therefore, the Lie derivative L_{X_t} satisfies, in analogy to Theorem B.34,

$$\frac{\mathrm{d}}{\mathrm{d}t} F_t^* = F_t^* L_{X_t} \, .$$

- As ω is closed, $L_{X_t}\omega = (d\,\mathbf{i}_{X_t} + \mathbf{i}_{X_t} d)\omega = d\,\mathbf{i}_{X_t}\omega$, and therefore

$$\frac{\mathrm{d}}{\mathrm{d}t} F_t^*\omega = F_t^* d\,\mathbf{i}_{X_t}\omega = d F_t^*\,\mathbf{i}_{X_t}\omega \, .$$

For $t_0 \in (0, 1]$, it follows by integration that

$$\omega - F_{t_0}^*\omega = F_1^*\omega - F_{t_0}^*\omega = \int_{t_0}^1 \frac{\mathrm{d}}{\mathrm{d}t} F_t^*\,\omega = d\int_{t_0}^1 F_t^*\,\mathbf{i}_{X_t}\,\omega\,\mathrm{d}t \, .$$

$F_{t_0}^*\omega$ will go to 0 in the limit $t_0 \searrow 0$, since $k \geq 1$. Therefore $\omega = d\varphi$ with $\varphi := \int_0^1 F_t^* \mathbf{i}_{X_t}\omega\,\mathrm{d}t$. □

We now will integrate 1-forms $\omega \in \Omega^1(U)$ along curves and see whether the result depends merely on the start and end points.

We assume without loss of generality that the open, non-empty subset $U \subseteq \mathbb{R}^n$ is *connected*, i.e., cannot be written as a disjoint union $U = U_1 \dot\cup U_2$ of such subsets.

B.46 Theorem *Let $U \subseteq \mathbb{R}^n$ be open. If $\omega \in \Omega^1(U)$ is a closed form, then*

$$\int_{c_0} \omega = \int_{c_1} \omega$$

for homotopic C^1-curves $c_0, c_1 : [a, b] \to U$ with common start and end points. (Refer to Definition A.22.)

Proof:

- By definition, there exists a *continuous* homotopy $h : [a, b] \times [0, 1] \to U$ from c_0 to c_1. Since U is open, this homotopy can be smoothed, for example by convolution.

So we assume that h is already smooth. We pull back the 1-form ω with h and obtain the 1-form $\tilde{\omega} := h^*\omega \in \Omega^1(\tilde{U})$ with[10] $\tilde{U} := [a, b] \times [0, 1]$.

- ω is closed, ($d\omega = 0$). By Theorem B.25, $\tilde{\omega}$ is thus closed as well.
- Now we write the boundary of the rectangle $\tilde{U}$ as the union of the images of the four curves

$$\begin{aligned} \tilde{c}_k : [a, b] \to \tilde{U} \ &, \ \tilde{c}_k(t) := (t, k) \qquad (k = 0, 1), \\ \tilde{a}, \tilde{b} : [0, 1] \to \tilde{U} \ &, \ \tilde{a}(s) := (a, s) \ , \ \tilde{b}(s) := (b, s). \end{aligned}$$

We have

$$\int_{\tilde{c}_k} \tilde{\omega} = \int_{c_k} \omega \quad , \quad \int_{\tilde{a}} \tilde{\omega} = \int_{\tilde{b}} \tilde{\omega} = 0 \,,$$

because $h \circ \tilde{c}_k = c_k, \ h \circ \tilde{a} \equiv a$, and $h \circ \tilde{b} \equiv b$.

- $\tilde{U}$ is convex, therefore the Poincaré lemma applies to the closed 1-form $\tilde{\omega}$. We choose the composite curves

$$d_0 : [a, b+1] \to \tilde{U} \ , \quad d_0(\tau) := \begin{cases} \tilde{c}_0(\tau) & , \ \tau \in [a, b] \\ \tilde{b}(\tilde{a}u - b) & , \ \tilde{a}u \in (b, b+1] \end{cases}$$

$$d_1 : [a, b+1] \to \tilde{U} \ , \ d_1(\tilde{a}u) := \begin{cases} \tilde{a}(\tilde{a}u) & , \ \tilde{a}u \in [0, 1] \\ \tilde{c}_1(\tilde{a}u - 1) & , \ \tilde{a}u \in (1, b+1] \end{cases}$$

They are piecewise smooth, with initial point $(a, 0)$ and end point $(b, 1)$, and

$$\int_{d_0} \tilde{\omega} = \int_{\tilde{c}_0} \tilde{\omega} + \int_{\tilde{b}} \tilde{\omega} = \int_{\tilde{c}_0} \tilde{\omega} \quad , \quad \int_{d_1} \tilde{\omega} = \int_{\tilde{a}} \tilde{\omega} + \int_{\tilde{c}_1} \tilde{\omega} = \int_{\tilde{c}_1} \tilde{\omega} \,.$$

By the Poincaré lemma, however, $\tilde{\omega} = d\tilde{\varphi}$; hence

$$\int_{d_k} \tilde{\omega} = \int_{d_k} d\varphi = \varphi(b, 1) - \varphi(a, 0) \qquad (k = 0, 1),$$

where in the last step, Stokes' theorem was used. It follows that

$$\int_{c_0} \omega = \int_{\tilde{c}_0} \tilde{\omega} = \int_{d_0} \tilde{\omega} = \int_{d_1} \tilde{\omega} = \int_{\tilde{c}_1} \tilde{\omega} = \int_{c_1} \omega \,. \qquad \square$$

B.47 Example (Irrotational Vector Fields)
If $v : U \to \mathbb{R}^n$ is a smooth vector field satisfying

$$\frac{\partial v_i}{\partial x_k} = \frac{\partial v_k}{\partial x_i} \qquad (i, k = 1, \ldots, n),$$

[10] It is not an issue that $\tilde{U}$ is not open, because we can extend h smoothly to an open superset of $\tilde{U}$.

then the 1-form $v^* \in \Omega^1(U)$ (namely $v^*(w) := \langle v, w\rangle$ for vector fields w) introduced in Section B.2 is closed, because

$$v^* = \sum_{k=1}^{n} v_k dx_k \quad \text{, hence} \quad dv^* = \sum_{i,k=1}^{n} \frac{\partial v_k}{\partial x_i} dx_i \wedge dx_k = 0\,.$$

If we integrate such vector fields along curves $c : [t_0, t_1] \to U$, by calculating

$$\int_c v^* = \int_{t_0}^{t_1} \left\langle v(c(t)), \frac{\mathrm{d}c}{\mathrm{d}t}(t) \right\rangle \mathrm{d}t\,,$$

then this integral is independent of the path, for homotopic curves. ◊

In the case of 1-forms in $\mathbb{R}^n$, the validity of Poincaré's lemma B.48 is not limited to star-shaped domains:

B.48 Theorem (Poincaré Lemma) *If the open set $U \subseteq \mathbb{R}^n$ is simply connected, closed 1-forms $\omega \in \Omega^1(U)$ are exact.*

Proof:

- As the open set $U \subseteq \mathbb{R}^n$ is connected, U is also *path connected*, i.e., from any point $x_0 \in U$, any other point $x \in U$ can be reached by a path $c_x : [0, 1] \to U$ with $c_x(0) = x_0$, $c_x(1) = x$. We may even assume c_x to be smooth.
- We define a function $\varphi : U \to \mathbb{R}$ by $\varphi(x) := \int_{c_x} \omega$. According to Theorem B.46, $\varphi(x)$ does not depend on the choice of c.
- φ is continuously differentiable, with $d\varphi = \omega$. □

B.49 Remark (Aharonov-Bohm) Example B.42 shows, in connection with Theorem B.48, that the punctured plane $\mathbb{R}^2 \backslash \{0\}$ is not simply connected. ◊

B.8 De-Rham Cohomology

Cohomology theories assign to a topological space certain groups that characterize the space to some extent. In this, they are similar to the homology theories (studied in Appendix G.2). De-Rham Cohomology uses differential forms and is thus Taylor-made for differentiable manifolds.

B.50 Definition *Let M be a differentiable manifold.*

- *For $k \in \mathbb{N}_0$ and the $\mathbb{R}$-vector space of closed k-forms*

$$Z^k(M) := \ker\big(d\restriction_{\Omega^k(M)}\big)\,,$$

two closed forms are called ***equivalent*** *or* ***cohomologous*** *($\omega_1 \sim \omega_2$) if $\omega_1 - \omega_2 \in B^k(M)$, with the subspace of exact forms*

$$B^k(M) := \mathrm{im}\big(d\restriction_{\Omega^{k-1}(M)}\big) \subseteq Z^k(M)\,.$$

- *The* k^{th} *de-Rham* ***cohomology group*** *is the quotient space*

$$H^k(M) := Z^k(M)/B^k(M) \quad \text{, and} \quad H^*(M) := \bigoplus_{k=0}^{\infty} H^k(M)\,. \tag{B.8.1}$$

For $k > \dim(M)$, these spaces have dimension zero.

For $k \le \dim(M)$, the $\mathbb{R}$-vector spaces $H^k(M)$ for compact manifolds M are, in contrast to $Z^k(M)$ and $B^k(M)$, finite dimensional; the same applies for non-compact manifolds that are not too complicated. So in this case, the **Betti numbers**

$$\mathrm{betti}_k(M) := \dim\big(H^k(M)\big) \qquad (k \in \mathbb{N}_0) \tag{B.8.2}$$

are defined. These numbers have topological relevance. For instance, one has

B.51 Theorem *If the manifold* M *is compact, the Betti number* $\mathrm{betti}_0(M)$ *is the number of connected components of* M.

Proof: A 0-form ω on M is a real-valued function.

- This form is closed if its derivative vanishes: $d\omega = 0$; which means it is constant (analogous to (B.6.2)) on the connected components of M.
- The values on the connected components can be chosen arbitrarily. □

B.52 Example (de-Rham Cohomology)

1. $M = \mathbb{R}^n$: $\omega \in \Omega^0(\mathbb{R}^n)$ is closed if and only if ω is constant. ω is exact if and only if this constant is 0 (compare Theorem B.51).
   In contrast, for $k \ge 1$, according to the Poincaré lemma (Theorem B.45), $\omega \in \Omega^k(\mathbb{R}^n)$ is closed if and only if ω is exact.
   Therefore $H^1(\mathbb{R}^n) = \ldots = H^n(\mathbb{R}^n) = \{0\}$ and $H^0(\mathbb{R}^n) = \mathbb{R}$.
2. **Circle** $M = S^1 = \mathbb{R}/\mathbb{Z}$:
   On $\mathbb{R}$, for $\lambda \in \mathbb{R}$, the 1-form $\tilde{\omega}_\lambda := \lambda\,dx$ is invariant under the deck transformations $x \mapsto x + \ell,\ \ell \in \mathbb{Z}$, and thus defines a 1-form ω_λ on S^1 with $\pi^*\omega_\lambda = \tilde{\omega}_\lambda$ for the projection $\pi : \mathbb{R} \to S^1,\ x \mapsto x + \mathbb{Z}$.
   This 1-form is closed, because there are no nontrivial forms of a higher degree than the dimension of the manifolds, but ω_λ is not exact, unless $\lambda = 0$. This can be seen from $\int_{S^1} \omega_\lambda = \lambda \neq 0$, so a function ρ_λ satisfying $d\rho_\lambda = \omega_\lambda =$ cannot exist (it would have to satisfy $\rho_\lambda(x) = \rho_\lambda(0) + \int_0^x \omega_\lambda$). In contrast, on $\mathbb{R}$, $\tilde{\omega}_\lambda = d\tilde{\rho}_\lambda$ with $\tilde{\rho}_\lambda(x) := \lambda x + c$, and $c \in \mathbb{R}$ arbitrary.
   Returning to S^1, every 1-form $\omega \in \omega^1(S^1)$ is *cohomologous* to a 1-form ω_λ, namely with $\lambda := \int_{S^1} \omega$:

$$\omega = \omega_\lambda + d\rho \quad \text{with} \quad \rho(x) := \int_0^x (\omega - \omega_\lambda) \qquad (x \in S^1).$$

Therefore, $H^0(S^1) = \mathbb{R}$ (again by Theorem B.51) and $H^1(S^1) = \mathbb{R}$. ◊

In many cases, when calculating de-Rham cohomologies, it is helpful to use that the exterior derivative behaves *naturally* under smooth mappings $f : M \to N$, namely that $df = fd$.

B.53 Example (Cohomology of Spheres) Calculation of the de-Rham cohomology $H^*(S^n)$ of the n-sphere is done by averaging differential forms.

- $S^n \subset \mathbb{R}^{n+1}$ is invariant under the operation of the rotation group $\mathrm{SO}(n+1)$, and we denote the restriction $g\restriction_{S^n}$ of a rotation $g : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ likewise as g.
- We can now average forms $\omega \in \Omega^*(S^n)$. We denote by μ the *Haar measure* on the group $\mathrm{SO}(n+1)$. It is the (unique) probability measure satisfying

$$\mu\big(L_g(A)\big) = \mu(A) = \mu\big(R_g(A)\big)$$

for all Borel sets $A \subseteq \mathrm{SO}(n+1)$, where left and right operation of $g \in \mathrm{SO}(n+1)$ are defined by

$$L_g(h) := g \circ h \quad , \quad R_g(h) := h \circ g \qquad \big(h \in \mathrm{SO}(n+1)\big)$$

(see also E.1.3). We average the form ω, i.e., we define

$$\overline{\omega} := \int_{\mathrm{SO}(n+1)} g^*\omega \, \mathrm{d}\mu(g)\,.$$

Then $h^*\overline{\omega} = \overline{\omega}$ for all $h \in \mathrm{SO}(n+1)$, so the form is invariant.
- Obviously, as $dg^* = g^*d$, the averaged from $\overline{\omega}$ is closed or exact if ω is closed or exact, respectively.
But the converse is also true: So we can pass to invariant representatives $\overline{\omega}$ of cohomology classes; representatives that are defined by their value at some point on the sphere, e.g., the north pole, and must be invariant under the stabilizer group $\mathrm{SO}(n) \subset \mathrm{SO}(n+1)$. This applies to constant 0-forms and n-forms, but for no other.
Therefore $H^0(S^n) = H^n(S^n) = \mathbb{R}$ and $H^i(S^n) = 0 \quad (0 < i < n)$. ◊

Sometimes it is also helpful that the exterior derivative for $\omega \in \Omega^k$ satisfies the Leibniz rule $d(\omega \wedge \rho) = (d\omega) \wedge \rho + (-1)^k \omega \wedge d\rho$. Therefore, if ω and ρ are closed, then so is $\omega \wedge \rho$, and for an arbitrary $(k-1)$-form σ, it is cohomologous to the closed form $(\omega + d\sigma) \wedge \rho$.

This makes $H^*(M)$ into a ring, the *cohomology ring*, and the *Künneth formula* holds:

$$H^k(M_1 \times M_2) = \bigoplus_{\ell_1+\ell_2=k} H^{\ell_1}(M_1) \otimes H^{\ell_2}(M_2) \qquad (k \geq 0)\,,$$

or briefly,

$$\boxed{H^*(M_1 \times M_2) = H^*(M_1) \otimes H^*(M_2)\,.}$$

B.54 Example (Cohomology of Tori)
For the n-torus $\mathbb{T}^n = S^1_1 \times \ldots \times S^1_n$, this implies that

$$H^k(\mathbb{T}^n) = \bigoplus_{\ell_1+\ldots+\ell_n=k} \bigotimes_{i=1}^{n} H^{\ell_i}(S^1_i) \cong \bigoplus_{\ell_1+\ldots+\ell_n=k} \mathbb{R} = \mathbb{R}^{\binom{n}{k}} \qquad (k = 0, \ldots, n).$$

Accordingly, the k^{th} Betti number is $\text{betti}_k(\mathbb{T}^n) = \binom{n}{k}$, and $H^*(\mathbb{T}^n)$ is the exterior algebra $\wedge(\omega_1, \ldots, \omega_n)$ generated by the ω_i from $H^*(S^1_i)$ $(i = 1, \ldots, n)$. $\Diamond$

Appendix C
Convexity and Legendre Transform

C.1 Convex Sets and Functions

C.1 Definition

- *A subset* $K \subseteq V$ *of a* $\mathbb{K}$*-vector space* V *is called* ***convex*** *if for every* $x, y \in K$*, the* ***segment between*** x ***and*** y*,*

$$[x, y] := \big\{(1-\lambda)x + \lambda y \mid \lambda \in [0,1]\big\} \subset V\,,$$

  lies in K*.*
- *A function* $f : K \to \mathbb{R} \cup \{+\infty\}$ *defined on a convex set* $K \subseteq V$ *is called* ***convex*** *if*

$$f\big((1-\lambda)x + \lambda y\big) \le (1-\lambda) f(x) + \lambda f(y) \qquad (\lambda \in [0,1]);$$

  $f : K \to \mathbb{R} \cup \{-\infty\}$ *is called* ***concave*** *if* $-f$ *is convex.*

C.2 Remark We have allowed the value ∞ here (with $-\infty < a < \infty$ for $a \in \mathbb{R}$ and the limited arithmetic $\infty + \infty := \infty$ and $a + \infty := \infty + a := \infty$ for $a \in \mathbb{R}$).

This can come in handy, because for a convex function $f : K \to \mathbb{R} \cup \{+\infty\}$, its extension

$$\tilde{f} : V \to \mathbb{R} \cup \{+\infty\} \quad , \quad \tilde{f}(x) := \begin{cases} f(x) & , \ x \in K \\ +\infty & , \ x \in V \setminus K \end{cases}$$

is still convex. Conversely, we can restrict functions $f : V \to \mathbb{R} \cup \{+\infty\}$ to their *effective domain*

$$\mathrm{dom}(f) := \{x \in V \mid f(x) \in \mathbb{R}\}\,,$$

which will be convex if f is convex. ◊

C.3 Example (Convexity)

1. All affine subspaces of V are convex.
2. Every norm on V is a convex function.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

3. For a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, the *epigraph of* f,

$$\text{epi}(f) := \left\{(x, t) \in \mathbb{R}^d \times \mathbb{R} \mid t \geq f(x)\right\},$$

is a convex subset of $\mathbb{R}^{d+1}$, if and only if f is convex.
4. For a convex function $f : V \to \mathbb{R} \cup \{+\infty\}$ and $c \in \mathbb{R} \cup \{+\infty\}$, the set $\{x \in V \mid f(x) \leq c\}$ is convex. $\Diamond$

Here are three facts about convex functions $f : K \to \mathbb{R}$ on *open* convex subsets $K \subseteq \mathbb{R}^d$ that are important to know:

- They are continuous.
- For all $x \in K \subseteq \mathbb{R}^d$, there exists (at least) one affine hyperplane in $\mathbb{R}^{d+1}$ that is the graph of an affine function

$$y \mapsto f(x) + \langle \mathcal{D}, y - x \rangle \qquad (y \in \mathbb{R}^d)$$

with

$$f(y) \geq f(x) + \langle \mathcal{D}, y - x \rangle \qquad (y \in K).$$

Such planes are called *supporting planes*, and if they are unique, *tangential planes*.
- If $f \in C^2(K, \mathbb{R})$, then f is convex if and only if the Hessian[11] $\mathrm{D}^2 f(x)$ is positive semidefinite at each point $x \in K$. In this case, the supporting plane is unique, and $\mathcal{D} = \mathrm{D}f(x)$.

C.2 The Legendre-Fenchel Transformation

C.4 Definition *The* ***Legendre-Fenchel transform*** *of a function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *(with* $f(\mathbb{R}^n) \neq \{+\infty\}$*) is*

$$f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \quad , \quad f^*(p) := \sup_{x \in \mathbb{R}^n} \big(\langle p, x \rangle - f(x)\big).$$

So in particular, $f^*(0) = -\inf_{x \in \mathbb{R}^n} f(x)$, and $f^*(p)$ is the vertical distance between the plane given by the graph of $x \mapsto \langle p, x \rangle$ and the graph of f.

Obviously, the *Young inequality*

$$\langle x, p \rangle \leq f(x) + f^*(p) \qquad (x, p \in \mathbb{R}^n)$$

[11]Recall: (1) The Hessian is of the form $\mathrm{D}^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$.
(2) A symmetric matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ is called *positive definite* ($A > 0$) (respectively *positive semidefinite*) if $\langle v, Av \rangle > 0$ (resp. $\langle v, Av \rangle \geq 0$) for $v \in \mathbb{R}^n \setminus \{0\}$.

follows. By allowing the value $+\infty$, it is certain that f^* exists.

C.5 Example (Legendre Transform of exp**)**
The exponential function $f(x) := e^x$ is defined on $\mathbb{R}$. The image $f'(\mathbb{R})$ is the interval $I := (0, \infty)$. For $p \in I$, the maximum $f^*(p) = \max_x(px - e^x)$ is attained at $x = \ln(p)$. So the Legendre transform of f is

$$f^* : I \to \mathbb{R}, \ f^*(p) = p\big(\ln(p) - 1\big),$$

see Figure C.2.1.

For the extended domain $\mathbb{R}$ of f^*, the supremum $f^*(0) = 0$ is no longer attained, and $f^*(p) = +\infty$ for $p < 0$. ◊

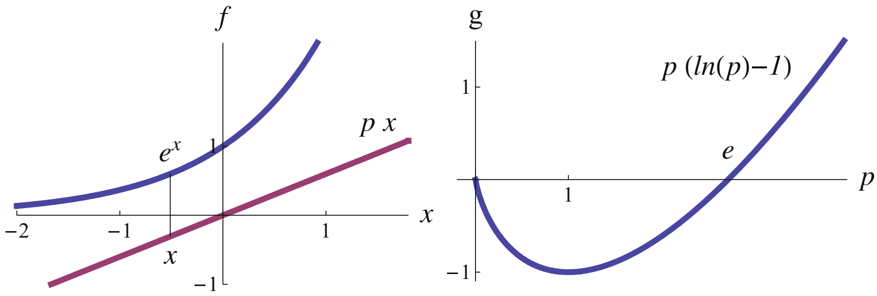


Figure C.2.1 Left figure pertaining to the definition of the Legendre transformation; right figure shows the Legendre transform $f^*(p) = p(\ln(p) - 1)$ of the exponential function

C.6 Theorem *The Legendre-Fenchel transform f^* is a convex function.*

Proof: For $p_0, p_1 \in \mathbb{R}^n$ and $p_t := (1-t)p_0 + tp_1$ $(t \in [0,1])$, we have

$$\begin{aligned}
f^*(p_t) &= \sup_{x \in \mathbb{R}^n} (\langle p_t, x\rangle - f(x)) \\
&= \sup_{x \in \mathbb{R}^n} \big((1-t)\big(\langle p_0, x\rangle - f(x)\big) + t\big(\langle p_1, x\rangle - f(x)\big)\big) \\
&\le (1-t) \sup_{x \in \mathbb{R}^n} \big(\langle p_0, x\rangle - f(x)\big) + t \sup_{x \in \mathbb{R}^n} \big(\langle p_1, x\rangle - f(x)\big) \\
&= (1-t) f^*(p_0) + t f^*(p_1) .
\end{aligned}$$
□

We now assume that f^* is not identically $+\infty$. Then $f^{**} := (f^*)^*$ is convex as well, and for all $y \in \mathbb{R}^n$, one has

$$\begin{aligned}
f^{**}(y) &= \sup_{p \in \mathbb{R}^n} \big(\langle y, p\rangle - f^*(p)\big) = \sup_{p \in \mathbb{R}^n} \inf_{x \in \mathbb{R}^n} \big(\langle y - x, p\rangle + f(x)\big) \\
&\le \sup_{p \in \mathbb{R}^n} \big(\langle y - y, p\rangle + f(y)\big) = f(y) .
\end{aligned}$$

So f^{**} is a convex function that is no bigger than f.

> From now on, we consider functions $f \in C^2(K, \mathbb{R})$ whose domain $K \subseteq \mathbb{R}^n$ is open and convex, extended by the value $+\infty$ to all of $\mathbb{R}^n$, and whose Hessian satisfies
> $$D^2 f(x) > 0 \quad (x \in K).$$

Defining $K^* := \Phi(K) \subseteq \mathbb{R}^n$ for $\Phi := Df$, one has

C.7 Theorem *Φ is a C^1-diffeomorphism from K onto K^*.*

Proof:

- By definition, $\Phi : K \to \Phi(K)$ is once continuously differentiable and surjective.
- But Φ is also injective, because for $x_0, x_1 \in K$, and with $x_t := (1-t)x_0 + tx_1$, one has

$$\langle \Phi(x_1) - \Phi(x_0), x_1 - x_0 \rangle = \int_0^1 \frac{d}{dt} \langle \Phi(x_t) - \Phi(x_0), x_1 - x_0 \rangle \, dt$$
$$= \int_0^1 \langle D\Phi(x_t)(x_1 - x_0), x_1 - x_0 \rangle \, dt > 0$$

  when $x_1 \neq x_0$, as a consequence of $D\Phi(x_t) = D^2 f(x_t) > 0$.
- As $D\Phi(x) > 0$, it follows from Theorem 2.39 that $\Phi(K)$ is open, and that the inverse of Φ is continuously differentiable, too. □

$\Phi(K)$ doesn't have to be convex. In any case, the following theorem holds:

C.8 Theorem $f^*(p) = \big\langle p, \Phi^{-1}(p) \big\rangle - f \circ \Phi^{-1}(p) \qquad \big(p \in \Phi(K)\big).$

Proof: For $p \in \Phi(K)$, we consider the function

$$g \in C^2(K, \mathbb{R}) \quad , \quad g(x) := \langle p, x \rangle - f(x)\,.$$

Then $D^2 g(x) = -D^2 f(x) < 0$, and for $x^* := \Phi^{-1}(p)$, one has

$$Dg(x^*) = p - Df(x^*) = p - \Phi \circ \Phi^{-1}(p) = 0\,.$$

By Taylor's formula for $x \in K$ and an appropriate y on the segment $[x^*, x]$, one has

$$g(x) - g(x^*) = \tfrac{1}{2} \big\langle x - x^*, D^2 g(y)(x - x^*) \big\rangle < 0\,,$$

provided $x \neq x^*$. So indeed, x^* is the unique location of a maximum of g. □

C.9 Theorem *1. If $f : K \to \mathbb{R}$ is twice continuously differentiable, then so is $f^* : K^* \to \mathbb{R}$.*

2. *If K^* is convex, then the Legendre transformation is an* ***involution****, i.e., $f^{**} = f$.*

Proof: 1. From $f^*(p) = \langle p, \Phi^{-1}(p)\rangle - f \circ \Phi^{-1}(p)$, it follows in view of $\Phi(x) = \mathrm{D}f(x)$ that

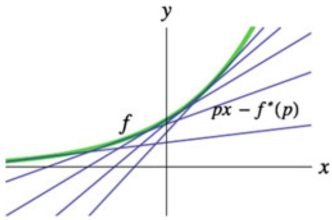
$$\mathrm{D}f^*(p) = \Phi^{-1}(p) + p^\top \mathrm{D}\Phi^{-1}(p) - \mathrm{D}f(\Phi^{-1}(p))\mathrm{D}\Phi^{-1}(p) = \Phi^{-1}(p)\,.$$

Now $\Phi^{-1} \in C^1(K^*, K)$, hence also $\mathrm{D}f^* \in C^1(K^*, K)$ and $f^* \in C^2(K^*, \mathbb{R})$.

2. Now for f^*, the same hypotheses hold as for f. Therefore $\mathrm{D}f^* : K^* \to K$ is the inverse mapping of $\mathrm{D}f : K \to K^*$, and by Thm. C.8, one has $f^{**} = f$. □

The Legendre transformation is an involution even for *arbitrary* (not necessarily differentiable) convex $f : K \to \mathbb{R}$, provided $K, K^* \subseteq \mathbb{R}^n$ are open and convex. For instance if K is an interval and $f \in C^1(K, \mathbb{R})$ is convex, then the graph of f lies above every tangent line to f.
This is because $f(x_t) \le (1-t)f(x_0) + tf(x_1)$ for $x_t := (1-t)x_0 + tx_1$, $t \in (0,1)$ implies $(f(x_t) - f(x_0))/t \le f(x_1) - f(x_0)$, hence

$$(x_1 - x_0)f'(x_0) = \lim_{t\to 0}\big(f(x_t) - f(x_0)\big)/t \le f(x_1) - f(x_0)\,.$$



f seen as an envelope of the family of lines defined by the Legendre transform f^*

C.10 Corollary *If $f \in C^1(K, \mathbb{R})$ is convex, then f is the upper envelope of the family*

$$y_p(x) = px - f^*(p) \qquad (p \in f'(K), x \in K)$$

of lines, see the figure to the right.

An analogous statement applies in the higher dimensional case for the tangential planes.

Appendix D
Fixed Point Theorems, and Results About Inverse Images

In the context of the Picard-Lindelöf theorem (Theorem 3.17) and also the Møller transformations (Theorem 12.11), there occur certain spaces of curves. The following theorem guarantees the convergence of Cauchy sequences in these spaces.

D.1 Theorem *Let $V \subseteq \mathbb{R}^n$ be closed, $I \subseteq \mathbb{R}$ an interval, and X the space of curves*

$$X := \left\{ c \in C(I, V) \;\middle|\; \sup_{t \in I} \|c(t)\| < \infty \right\},$$

with $\;d(f, g) := \sup_{t\in I} \|f(t) - g(t)\| \quad (f, g \in X)$.

Then (X, d) is a complete metric space.

Proof: • (X, d) is a metric space. Indeed, for $f, g, h \in X$, one has

- $d(f, g) \leq \sup_{t\in I} \|f(t)\| + \sup_{t\in I} \|g(t)\| < \infty$, and $d(f, g) \geq 0$, with equality if and only if $f = g$,
- $d(g, f) = d(f, g)$, and
- $d(f, h) = \sup_{t\in I} \|(f(t) - g(t)) + (g(t) - h(t))\|$
  $\leq \sup_{t\in I} \|f(t) - g(t)\| + \sup_{t\in I} \|g(t) - h(t)\| = d(f, g) + d(g, h)$.

- Cauchy sequences $(f_m)_{m\in\mathbb{N}}$ in X converge pointwise to a mapping $f : I \to V$, because for all $t \in I$, one has $\|f_m(t) - f_n(t)\| \leq d(f_m, f_n)$, and V is complete, being a closed subset of $\mathbb{R}^n$.
- The following $\varepsilon/3$-*argument* shows that f is continuous, and even $f \in X$. For $\varepsilon > 0$, let $N \in \mathbb{N}$ be chosen such that $d(f_m, f_N) < \varepsilon/3 \quad (m \geq N)$, and thus also for all $t \in I : \; \|f(t) - f_N(t)\| \leq \varepsilon/3$.
  From the continuity of f_N, it follows for all $s \in I$ that there exists a $\delta > 0$ for which
  $$\|f_N(t) - f_N(s)\| < \tfrac{\varepsilon}{3} \qquad \big(t \in I : |t - s| < \delta/3\big).$$
  Therefore,
  $$\|f(t) - f(s)\| \leq \|f(t) - f_N(t)\| + \|f_N(t) - f_N(s)\| + \|f_N(s) - f(s)\| < \varepsilon\,.$$

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

- In (X, d), one also has $f = \lim_{N\to\infty} f_N$, since the distance

$$d(f, f_N) = \sup_{t\in I} \lim_{m\to\infty} \|f_m(t) - f_N(t)\| \le \sup_{t\in I} \sup_{m\ge N} \|f_m(t) - f_N(t)\|$$

$= \sup_{m\ge N} d(f_m, f_N)$ converges to 0 as $N \to \infty$. □

D.2 Definition • *$x \in X$ is called a* ***fixed point*** *of $f : X \to X$ if $f(x) = x$.*
- *If (X, d) is a complete metric space, then $f : X \to X$ is called a* ***contraction*** *if there exists a* ***contraction constant*** *$\theta \in (0, 1)$ such that*

$$d\big(f(x), f(y)\big) \le \theta\, d(x, y) \qquad (x, y \in X).$$

D.3 Theorem (Banach's Fixed Point Theorem) *A contraction $f : X \to X$ on a complete metric space (X, d) with Lipschitz constant $\theta < 1$ has exactly one fixed point x^*, and the m^{th}* ***iterate*** *$x_m := f(x_{m-1})$* ***of*** *$x_0 \in X$ satisfies*

$$\boxed{d(x_m, x^*) \le d(x_1, x_0)\frac{\theta^m}{1-\theta} \qquad (m \in \mathbb{N}).} \tag{D.1}$$

Proof: • For $n \in \mathbb{N}_0$, we have $d(x_{n+1}, x_n) \le \theta\, d(x_n, x_{n-1}) \le \ldots \le \theta^n d(x_1, x_0)$, hence for $n > m$,

$$d(x_n, x_m) \le \sum_{j=m}^{n-1} d(x_{j+1}, x_j) \le \Big(\textstyle\sum_{j=m}^{n-1} \theta^j\Big)\, d(x_1, x_0) \le \frac{\theta^m}{1-\theta}\, d(x_1, x_0)\,.$$

So $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, and by completeness of (X, d), there exists

$$x^* := \lim_{n\to\infty} x_n\,.$$

- Therefore, $d(x^*, x_m) = \lim_{n\to\infty}\big(d(x^*, x_n) + d(x_n, x_m)\big) \le \frac{\theta^m}{1-\theta}\, d(x_1, x_0)$.
- Continuity of f implies $f(x^*) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x^*$.
- If some $x \in X$ also satisfies $f(x) = x$, then $d(x, x^*) = d\big(f(x), f(x^*)\big) \le \theta\, d(x, x^*)$, hence $x = x^*$. So x^* is the only fixed point. □

In many cases, the contraction depends on a parameter and one is interested in how the fixed point depends on this parameter. A theorem to this effect is:

D.4 Theorem (Parametrized Fixed Point Theorem)
Let (X, d) be a complete metric space, P a topological space (the ***parameter space****) and let $f : X \times P \to X,\; f_p(x) := f(x, p)$ satisfy:*

1. *There exists a common contraction constant $\theta \in (0, 1)$ for the mappings $f_p \quad (p \in P)$.*
2. *For all $x \in X$, the mappings $P \to X,\; p \mapsto f_p(x)$ are continuous.*

Then the fixed points x_p of f_p are continuous in $p \in P$.

Proof: By #1, f_p has a unique fixed point x_p according to Theorem D.3.

- For all $p, q \in P$, we have, again due to the first hypothesis,

$$\begin{aligned} d(x_p, x_q) &= d\big(f_p(x_p), f_q(x_q)\big) \le d\big(f_p(x_p), f_p(x_q)\big) + d\big(f_p(x_q), f_q(x_q)\big) \\ &\le \theta\, d\big(x_p, x_q\big) + d\big(f_p(x_q), f_q(x_q)\big)\,, \end{aligned}$$

  hence $d(x_p, x_q) \le d\big(f_p(x_q), f_q(x_q)\big)/(1-\theta)$.
- Due to the first condition, for all $\varepsilon > 0$, there exists a neighborhood $U \subseteq P$ of q with $d\big(f_p(x_q), f_q(x_q)\big) < \varepsilon(1-\theta)$, if $p \in U$. For these p, one therefore has $d(x_p, x_q) < \varepsilon$. □

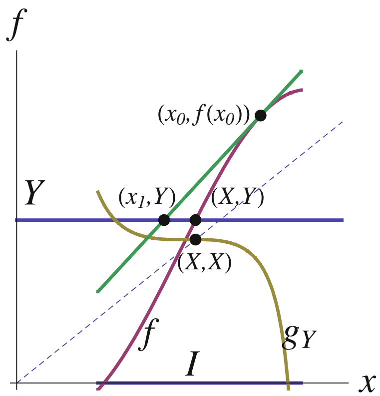
D.5 Remark (Newton Method) Let $f \in C^2(I, \mathbb{R})$, i.e., a twice continuously differentiable function on the interval $I := [a, b]$, with $f'(x) \neq 0$ for all $x \in I$.

We want to find the pre-image X (which is unique as a consequence of that hypothesis) of a point[12] $Y \in f(I)$. X is the only fixed point of

$$g \equiv g_Y : I \to \mathbb{R} \quad , \quad g(x) = x - \tfrac{f(x)-Y}{f'(x)}.$$

Geometrically, $g(x)$ is the intersection of the tangent to the graph of f at $(x, f(x))$ with the line $y = Y$.

In the Newton method, after choosing an initial value $x_0 \in I$, one iterates g, i.e., $x_{n+1} := g(x_n)$. ◊



D.6 Theorem (Newton Method) *For all $x \in I$, one has the estimate $|g(x) - X| \le M\,|x - X|^2$, where $M := \frac{\max_{x\in I}|f''(x)|}{\min_{x\in I}|f'(x)|}$. Therefore, the method converges for initial values $x_0 \in I$ satisfying $M|x_0 - X| < 1$, i.e.: $\lim_{n\to\infty} x_n = X$.*

Proof: If $x = X$, we are done already. Otherwise, by the mean value theorem for derivatives, there exists ξ in the open interval between x and X with $\frac{f(x)-f(X)}{x-X} = f'(\xi)$, and therefore

$$\begin{aligned} |g(x) - X| &= \left| x - \tfrac{f(x)-Y}{f'(x)} - X \right| = \left| (x - X)\left(1 - \tfrac{f'(\xi)}{f'(x)}\right)\right| \\ &= |x - X| \cdot \left| \tfrac{f'(x)-f'(\xi)}{f'(x)} \right| \le |x - X|^2 \cdot \left| \tfrac{f'(x)-f'(\xi)}{x-\xi} \tfrac{1}{f'(x)} \right| \\ &= |x - X|^2 \cdot \tfrac{|f''(\eta)|}{|f'(x)|}\,. \end{aligned}$$

Again we have chosen η appropriately between x and ξ according to the mean value theorem. □

[12] Frequently, one assumes $Y = 0$ in discussions of Newton's method.

Appendix E
Group Theory

E.1 Groups

E.1 Definition

- *A* ***group*** *is a set G with a mapping (called 'operation')*

$$\circ : G \times G \to G$$

  that is associative $(g \circ (h \circ k) = (g \circ h) \circ k)$*, and has a distinguished element* $e \in G$ ***(identity)*** *with* $g \circ e = e \circ g = g$ $(g \in G)$ *such that for every element* $g \in G$*, there exists an* ***inverse element*** $h \in G$ *with* $g \circ h = e$.
- *G is called* ***commutative*** *or* ***abelian*** *if* $g \circ k = k \circ g \quad (g, k \in G)$.
- *The* ***order*** $|G|$ *of a group is the cardinality of the set G.*
- *A nonempty subset* $H \subseteq G$ *is called a* ***subgroup*** *if it is closed under multiplication and inverses. We write:* $H \leq G$.
- *For a subgroup* $H \leq G$ *and* $g \in G$*, we call* $g \circ H := \{g \circ h \mid h \in H\} \subseteq G$ *a* ***left coset****, and* $H \circ g$ *a* ***right coset*** *of H.*

E.2 Remark Both the identity e and the inverse g^{-1} of g is unique, and one also has $g^{-1} \circ g = e$. ◊

E.3 Theorem (Lagrange) *The order* $|H|$ *of a subgroup* $H \leq G$ *of a finite group G divides the order* $|G|$ *of G.*

E.4 Definition $g_1 \in G$ *is called* ***conjugate*** *to* $g_2 \in G$ *if there exists an* $h \in G$ *with* $g_2 = h \circ g_1 \circ h^{-1}$*. A subgroup* $H_1 \leq G$ *is called* ***conjugate*** *to* $H_2 \leq G$ *if there exist* $h \in G$ *with* $H_2 = h \circ H_1 \circ h^{-1}$.

E.5 Theorem *Conjugacy is an equivalence relation. A group G is therefore decomposed into conjugacy classes, where the identity* $e \in G$ *is a class of its own.*

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

E.6 Definition *A subgroup $H \leq G$ is called **normal** if*

$$kHk^{-1} = H \qquad (k \in G),$$

in other words, if left and right cosets coincide. We write: $H \lhd G$.

E.7 Theorem
*If the subgroup $H \leq G$ is normal, the operation on G induces an operation on the set G/H of cosets of H. This group is called the **factor group** G/H.*

E.8 Example (Group of Residues)
For $n \in \mathbb{N}$, the subset $n\mathbb{Z} \subseteq \mathbb{Z}$ of integer multiples of n is a subgroup of $(\mathbb{Z}, +)$. As $(\mathbb{Z}, +)$ is abelian, $n\mathbb{Z}$ is normal, and the factor group $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z} \cong \{0, 1, \dots, n - 1\}$ is called the *group of residues* (with addition mod n). ◊

E.9 Definition *A mapping $\phi : G_1 \to G_2$ from the group G_1 into the group G_2 is called a **homomorphism** if*

$$\phi(g_1 \circ g_2) = \phi(g_1) \circ \phi(g_2) \qquad (g_1, g_2 \in G_1). \tag{E.1.1}$$

*A bijective homomorphism is called an **isomorphism**.*

E.10 Remark (Group Homomorphisms)
From (E.1.1), it follows for the images of the neutral and inverse elements that

$$\phi(e) = e \quad \text{and} \quad \phi(g^{-1}) = \phi(g)^{-1} \quad (g \in G_1). \qquad \Diamond$$

E.11 Theorem *The **kernel** $\ker(\phi) := \{g \in G_1 \mid \phi(g) = e\}$ of a group homomorphism $\phi : G_1 \to G_2$ is a normal subgroup.*

E.12 Definition

- *A **(left) operation**, or **action** of a group G on a set M is a mapping $\Phi : G \times M \to M$ which, in the notation*

$$\Phi_g : M \to M \quad , \quad \Phi_g(m) := \Phi(g, m) \qquad (g \in G)\,,$$

satisfies

$$\Phi_e = Id_M \quad \textit{and} \quad \Phi_g \circ \Phi_h = \Phi_{g \circ h} \qquad (g, h \in G). \tag{E.1.2}$$

- *The **orbit** of a point $m \in M$ is the set $\mathcal{O}(m) := \Phi(G, m) \subseteq M$.*
- *The **isotropy group** (also called **stabilizer**) of $m \in M$ is the subgroup*

$$G_m := \{g \in G \mid \Phi(g, m) = m\}\,.$$

- *The action is called **free** if for all $m \in M$, the mapping $G \to M,\ g \mapsto \Phi(g, m)$ is injective;*

it is called ***transitive*** *if this mapping is surjective for some* $m \in M$ *(and thus automatically for all* $m \in M$*).*

- *A group action* $\Phi : G \times V \to V$ *on a vector space* V *is called a* ***representation*** *of* G *if the mappings* $\Phi_g : V \to V$ *are linear.*
  $\Phi : G \times \mathbb{C} \to \mathbb{C}$, $(g, c) \mapsto \varphi(g)c$ *(and also* $\varphi : G \to \mathbb{C}$*) is called a* ***character****.*

Just as in the special case of dynamical systems (Theorem 2.13), membership in an orbit defines an equivalence relation on M. Let's use the shorthand notation $g\,m := \Phi_g(m)$. Then, for $h \in G$ and the point $m' := h\,m \in \mathcal{O}(m)$ of the orbit of m, the isotropy groups satisfy

$$G_{m'} = \{g \in G \mid g \circ h\,m = h\,m\} = \{g \in G \mid h^{-1} \circ g \circ h\,m = m\} = h\,G_m\,h^{-1}\,.$$

The isotropy groups of the points of an orbit are therefore conjugate to each other.

Groups G operate on themselves in various ways, for instance by *left* or *right action*[13] respectively:

$$L : G \times G \to G \quad , \quad L_g(h) := g \circ h \quad , \quad R : G \times G \to G \quad , \quad R_g(h) := h \circ g\,. \tag{E.1.3}$$

The mappings L_g and $R_{g'}$ commute, and both left and right action are transitive, because given $h_1, h_2 \in G$, one has $L_g(h_1) = h_2$ for $g := h_2 \circ h_1^{-1}$ and $R_g(h_1) = h_2$ for $g := h_1^{-1} \circ h_2$. The mapping

$$I : G \times G \to G \quad , \quad I_g(h) := g \circ h \circ g^{-1} \tag{E.1.4}$$

is also a group action of G on itself, but in contrast to L_g and R_g, $I_g : G \to G$ is a group automorphism, because

$$I_g(h_1) \circ I_g(h_2) = g \circ h_1 \circ g^{-1} \circ g \circ h_2 \circ g^{-1} = I_g(h_1 \circ h_2)\,,$$

so in particular, $I_g(e) = e$ for all $g \in G$. This group action is called *conjugation* (according to Definition E.4).

E.13 Theorem *If* $|G| < \infty$ *and* $\Phi : G \times M \to M$ *is a group action, then*

$$|\mathcal{O}(m)| \cdot |G_m| = |G| \qquad (m \in M).$$

E.14 Definition

- *If* G_1, G_2 *are groups, then* $G_1 \times G_2$ *together with the group multiplication*

$$(g_1, g_2) \circ (g_1', g_2') := (g_1 \circ g_1', g_2 \circ g_2')$$

  is called the ***direct product*** *of* G_1 *and* G_2*.*

[13] However, the right action satisfies $R_{g_1} \circ R_{g_2} = R_{g_2 \circ g_1}$, in contrast to (E.1.2).

- *The **automorphism group** of a group G is the group*

$$\mathrm{Aut}(G) := \{\phi : G \to G \mid \phi \textit{ is an isomorphism }\}$$

with composition as the group operation.
- *If $\phi : G_2 \to \mathrm{Aut}(G_1)$ is a homomorphism of G_2 into the automorphism group of G_1, then we call the set $G_1 \times G_2$ together with the group multiplication*

$$(g_1, g_2) \circ (g_1', g_2') = \big(g_1 \circ \big(\phi(g_2)(g_1')\big)\,,\; g_2 \circ g_2'\big) \qquad \big(g_i, g_i' \in G_i\big)$$

*a **semidirect product** of G_1 and G_2; we denote it as $G_1 \rtimes_\phi G_2$ (or $G_1 \rtimes G_2$).*

E.15 Example (Euclidean Group)
As is shown in Theorem 14.1, the Euclidean group $\mathbb{E}(d)$ of $\mathbb{R}^d$ is the semidirect product of $G_1 := (\mathbb{R}^d, +)$ and the orthogonal group $G_2 := \mathrm{O}(d)$. Hereby, the homomorphism $\phi : \mathrm{O}(d) \to \mathrm{Aut}(\mathbb{R}^d)$ is simply given by $\phi(O)(v) := Ov$ for $v \in \mathbb{R}^d$, so that

$$(v, O) \circ (v', O') = \big(v + Ov', O\, O'\big)\,.$$

In particular for $d = 1$ dimension, one obtains $\mathbb{E}(1) \cong \mathbb{R} \rtimes_\phi \{\pm 1\}$, where this semidirect product of abelian groups is not abelian; indeed, for example,

$$(v, -1) \circ (v', 1) = (v - v', -1) \quad \text{, but} \quad (v', 1) \circ (v, -1) = (v + v', -1)\,. \qquad \diamondsuit$$

Whereas both factors of a semidirect product are subgroups of $G_1 \rtimes G_2$, in general, only G_1 is normal.
For instance, in the example of the Euclidean group, the subgroup $\{0\} \times \mathrm{O}(d)$ of rotations and rotoreflections about the origin is conjugate to the subgroup of rotations and rotoreflections about another point of $\mathbb{R}^d$.

E.2 Lie Groups

Manygroups are topological spaces or even manifolds in a natural way.

E.16 Definition

- *A group $(G, \circ)$ is called a **topological group** if G is equipped with a topology such that:*
  The group operation $\circ : G \times G \to G$ and the inverse $G \to G$ are continuous (here $G \times G$ carries the product topology, see page 488).
- *A group $(G, \circ)$ is called a **Lie group** if G carries a differentiable structure (see page 491) such that the group operation $G \times G \to G$ and the inverse $G \to G$ are smooth mappings.*

A simple example of a Lie group is the abelian group $(\mathbb{R}^d, +)$, and also its subgroup $\mathbb{Z}^d$ with the subspace topology from the inclusion $\mathbb{Z}^d \subset \mathbb{R}^d$ (namely the discrete topology).

E.17 Remark (Topological Groups and Lie Groups)

1. We already made use of Definition E.16 when we defined a continuous dynamical system (Definition 2.16) and a differentiable dynamical system (Definition 2.43).
2. As manifolds are in particular topological spaces, and smooth mappings are continuous, every Lie group is a topological group.
   The converse is not true. An example is the group $\mathrm{Diff}(M)$ of diffeomorphisms of a manifold M; it can be viewed as a topological group. When $\dim(M) \geq 1$, this group would have to have infinite dimension as a 'manifold', which is however not compatible with our definition of a manifold.
3. To show that some G is a Lie group, it suffices to show that the multiplication is a smooth operation. Indeed, if we denote it as $M : G \times G \to G$, then taking the inverse $I : G \to G,\ I(g) = g^{-1}$ amounts to solving the equation $M\big(g, I(g)\big) = e$, and the partial derivative $\mathrm{D}_2 M(g, h)$ with respect to the second argument is an isomorphism. So the smoothness of I follows from the implicit function theorem. ◊

E.18 Example (General Linear Group)
The *general linear group* $\mathrm{GL}(V)$ of a vector space V is the group of automorphisms of V (i.e., of the *invertible* linear mappings in $\mathrm{Lin}(V)$).

So it operates on V and also on the *projective space* $\mathrm{P}(V)$ of V, which consists of the set of equivalence classes

$$\mathrm{P}(V) := \big\{[v] \mid v \in V \setminus \{0\}\big\} \quad \text{with} \quad [v] = [w] \text{ iff } \mathrm{span}(v) = \mathrm{span}(w)\,. \tag{E.2.1}$$

In the case of a $\mathbb{K}$-vector space V of finite dimension n, this group is isomorphic to the group of all invertible matrices from $\mathrm{Mat}(n, \mathbb{K})$. For $\mathbb{K} = \mathbb{R}$ and all $n \in \mathbb{N}$, this *real general linear group*

$$\mathrm{GL}(n, \mathbb{R}) := \{M \in \mathrm{Mat}(n, \mathbb{R}) \mid \det(M) \neq 0\}$$

is an open subset of the n^2-dimensional vector space $\mathrm{Mat}(n, \mathbb{R})$, and the group multiplication is smooth, since it is just the restriction of the bilinear matrix multiplication (see also Example 4.13). So $\dim\big(\mathrm{GL}(n, \mathbb{R})\big) = n^2$. The Lie group $\mathrm{GL}(n, \mathbb{R})$ has, next to the subgroup

$$\mathrm{GL}^+(n, \mathbb{R}) := \{M \in \mathrm{GL}(n, \mathbb{R}) \mid \det(M) > 0\}\,,$$

one more connected component. This latter is of the form $\mathrm{GL}^+(n, \mathbb{R})N$ for any arbitrary matrix N with negative determinant. As a manifold, it therefore looks exactly like $\mathrm{GL}^+(n, \mathbb{R})$, but it is not a group.

A subgroup of $\mathrm{GL}(n, \mathbb{R})$ that is of interest is the *orthogonal group*

$$\mathrm{O}(n) := \left\{ M \in \mathrm{GL}(n,\mathbb{R}) \mid M^\top = M^{-1} \right\} ,$$

which consists of rotations and rotoreflections of $\mathbb{R}^n$. The topology of the manifold $\mathrm{GL}(n,\mathbb{R})$ can be understood by writing matrices $M \in \mathrm{GL}(n,\mathbb{R})$ uniquely in the form $M = OP$ with $O \in \mathrm{O}(n)$ and P positive; this is the *polar decomposition*. It is clear from the ansatz $M = OP$ that P has to equal $(M^\top M)^{1/2}$: indeed, $M^\top M = P^\top O^\top O P = P^\top P = P^2$. Since positive matrices are of the form $P = \exp(S)$ with $S \in \mathrm{Sym}(n,\mathbb{R})$,[14] the mapping

$$\mathrm{GL}(n,\mathbb{R}) \longrightarrow \mathrm{O}(n) \quad , \quad OP \longmapsto O$$

gives a bundle with the typical fiber being $\mathrm{Sym}(n,\mathbb{R})$. As an $\mathbb{R}$-vector space, this fiber is connected. Therefore the total space $\mathrm{GL}(n,\mathbb{R})$ of the bundle has exactly as many connected components as the base space $\mathrm{O}(n)$. Now $\mathrm{GL}^+(n,\mathbb{R})$ projects to the rotation group

$$\mathrm{SO}(n) := \left\{ M \in \mathrm{GL}^+(n,\mathbb{R}) \mid M^\top = M^{-1} \right\}.$$

This latter is connected, as is shown in Example E.19.

On the other hand, the group $\mathrm{GL}(n,\mathbb{R})$ is *not compact*, since it contains the subgroup $\{\lambda \mathbb{1} \mid \lambda > 0\}$.

Analogously, for all $n \in \mathbb{N}$, the *complex general linear group*

$$\mathrm{GL}(n,\mathbb{C}) := \{M \in \mathrm{Mat}(n,\mathbb{C}) \mid \det M \neq 0\}$$

is a $2n^2$-dimensional (real) manifold, being an open dense subset of $\mathrm{Mat}(n,\mathbb{C})$. This Lie group however is connected. ♢

Many Lie groups are subgroups of $\mathrm{GL}(n,\mathbb{R})$ or $\mathrm{GL}(n,\mathbb{C})$.

E.19 Example (Lie Groups)

1. The *special linear group*

   $$\mathrm{SL}(n,\mathbb{R}) := \{M \in \mathrm{GL}(n,\mathbb{R}) \mid \det(M) = 1\}$$

   is an $(n^2 - 1)$-dimensional submanifold of $\mathrm{GL}^+(n,\mathbb{R})$ since 1 is a regular value of $\det : \mathrm{Mat}(n,\mathbb{R}) \to \mathbb{R}$.
   $\mathrm{SL}(n,\mathbb{R})$ is connected, and for $n > 1$, it is not compact.
2. The *rotation group* $\mathrm{SO}(n) = \left\{ M \in \mathrm{SL}(n,\mathbb{R}) \mid M^\top = M^{-1} \right\}$ equals the pre-image of 1 for $\det : \mathrm{O}(n) \to \{-1, 1\}$.
   It is the connected component of the orthogonal group $\mathrm{O}(n)$ that contains the identity. This can be seen using the real Jordan normal form, since any matrix $O \in \mathrm{SO}(n)$, being a normal matrix, is, with respect to an appropriate orthonormal

[14] Here $\mathrm{Sym}(n,\mathbb{K}) := \{P \in \mathrm{GL}(n,\mathbb{K}) \mid P^* = P\}$ for $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

basis of $\mathbb{R}^n$, a direct sum (block diagonal matrix) of matrices of the form $\left(\begin{smallmatrix} c & -s \\ s & c \end{smallmatrix}\right) \in \mathrm{Mat}(2,\mathbb{R})$ and numbers $u \in \mathrm{Mat}(1,\mathbb{R}) = \mathbb{R}$. Since O is orthogonal, the matrices $\left(\begin{smallmatrix} c & -s \\ s & c \end{smallmatrix}\right)$ are $\in \mathrm{SO}(2)$, and $|u| = 1$. Since moreover $O \in \mathrm{SO}(n)$, there is an even number of terms in the sum with $u = -1$. These can be combined in pairs to rotation matrices in SO(2). Since $\mathrm{SO}(2) = \left\{\left(\begin{smallmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{smallmatrix}\right) \mid \varphi \in [0, 2\pi)\right\}$ is connected, so is $\mathrm{SO}(n)$.
As the matrix norm of elements of $\mathrm{O}(n)$ equals 1, $\mathrm{O}(n)$, and thus also $\mathrm{SO}(n)$, is *compact*.
$1\!\!1 \in \mathrm{Sym}(n,\mathbb{R}) \subset \mathrm{Mat}(n,\mathbb{R})$ is a regular value of the mapping

$$P : \mathrm{GL}(n,\mathbb{R}) \longrightarrow \mathrm{Sym}(n,\mathbb{R}) \quad , \quad M \longmapsto M^\top M \,,$$

because the linearization is given by $\mathrm{D}P(M)N = M^\top N + N^\top M$, and for $M \in P^{-1}(1\!\!1) = \mathrm{O}(n)$ and $L \in \mathrm{Sym}(n,\mathbb{R})$, we see that L is the image of $N := \frac{1}{2}ML$ under the mapping $N \mapsto M^\top N + N^\top M$, hence the linearization is surjective. Thus $\mathrm{O}(n)$ is a *Lie group*, and

$$\dim\bigl(\mathrm{O}(n)\bigr) = \dim\bigl(\mathrm{GL}(n,\mathbb{R})\bigr) - \dim\bigl(\mathrm{Sym}(n,\mathbb{R})\bigr) = n^2 - \binom{n+1}{2} = \binom{n}{2} \,.$$

3. The *unitary group*

$$\mathrm{U}(n) := \left\{M \in \mathrm{GL}(n,\mathbb{C}) \mid M^* = M^{-1}\right\}$$

is also a compact Lie group; it has the (real) dimension n^2. Specifically $U(1) \cong S^1$. The subgroup

$$\mathrm{SU}(n) := \{M \in \mathrm{U}(n) \mid \det(M) = 1\}$$

of *special unitary* matrices is an $(n^2 - 1)$-dimensional Lie group. The representation

$$\mathrm{SU}(2) = \left\{\left(\begin{smallmatrix} v & w \\ -\overline{w} & \overline{v} \end{smallmatrix}\right) \mid v, w \in \mathbb{C}, |v|^2 + |w|^2 = 1\right\} \qquad \text{(E.2.2)}$$

shows that SU(2) is diffeomorphic to the sphere $S^3 \subset \mathbb{R}^4 \cong \mathbb{C}^2$.
4. *Cartesian products* of Lie groups are again Lie groups, with the manifold understood as the product manifold. An example of this is the abelian group $\mathbb{T}^n := S^1 \times \ldots \times S^1$, which is the *n-dimensional torus*. ◊

E.3 Lie Algebras

The concept of *linearization* is frequently successful in mathematics, because linear structures are usually easier to understand than nonlinear ones. As a Lie group is in particular a manifold, one can study it locally in the neighborhood of the neutral element, and in doing so, one is led to the notion of a Lie algebra.

In the case of a Lie group G, the left and right actions $L_g,\ R_g : G \to G$ defined in (E.1.3) are diffeomorphisms.

E.20 Definition *A vector field $X : G \to TG$ on a Lie group G is called*

- ***left invariant*** *if* $(L_g)_* X = X \quad (g \in G)$,
- ***right invariant*** *if* $(R_g)_* X = X \quad (g \in G)$.

We will subsequently consider mainly left invariant vector fields X; for right invariant vector fields, analogous statements apply.

As the left action is transitive, it suffices to know $X(e)$; then $X(g) = TL_g X(e)$ is determined for all $g \in G$. Thus the vector space of left invariant vector fields on G, which is a subspace $\mathcal{X}_L(G)$ of the $\mathbb{R}$-vector space $\mathcal{X}(G)$ of all vector fields, is naturally isomorphic to T_eG.

E.21 Definition *A **Lie algebra** $\big(E, [\cdot,\cdot]\big)$ is a K-vector space E with a mapping $[\cdot,\cdot] : E \times E \to E$, called the **Lie bracket**, that satisfies the following properties (where $a, b \in K$ and $X, Y, Z \in E$):*

- *It is **bilinear**, i.e.,*

$$[a\,X + b\,Y, Z] = a\,[X, Z] + b\,[Y, Z] \quad \text{and} \quad [Z, a\,X + b\,Y] = a\,[Z, X] + b\,[Z, Y]$$

- *It is **alternating**: $[X, X] = 0$, and therefore*
  ***antisymmetric**: $[X, Y] = -[Y, X]$,*
- *It satisfies the **Jacobi identity** $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$.*

E.22 Lemma *The Lie bracket $[X, Y] : G \to TG$ (see Definition 10.20) of two left invariant vector fields $X, Y : G \to TG$ is again left invariant.*

Proof: For $g \in G$, one has $(L_g)_*[X, Y] = [(L_g)_* X, (L_g)_* Y] = [X, Y]$, hence $[X, Y] \in \mathcal{X}_L(G)$. □

In consequence, $\big(\mathcal{X}_L(G), [\cdot,\cdot]\big)$ is a Lie algebra; it is called the *Lie algebra* $\mathrm{Lie}(G)$ *of* G.

E.23 Remark (Lie Algebras)

1. One frequently uses the gothic font symbol $\mathfrak{g}$ instead of $\mathrm{Lie}(G)$. For instance

$$\mathrm{Lie}\big(\mathrm{SO}(n)\big) = \mathfrak{so}(n)\,.$$

2. As noted, one can view the Lie algebra $\mathfrak{g}$ of a Lie group G as the tangent space of G at the neutral element $e \in G$. ([15]) As such, it describes the local structure of the

[15] The *Baker-Campbell-Hausdorff formula* expresses $\log\big(\exp(X)\exp(Y)\big)$ in terms of commutators of X and Y, where exp is taken from (E.3.1); so it gives a relation between the Lie bracket and the group operation.

group near the identity, and it is therefore possible, as in Exercise E.27.2.b below, or the example $\mathfrak{so}(n) = \mathfrak{o}(n)$, that non-isomorphic Lie groups have isomorphic Lie algebras. ◊

Not all Lie groups are matrix groups:

E.24 Example (Lie Groups vs. Matrix Groups)

1. Whereas $(\mathbb{R}, +)$ consists of real numbers, i.e., one dimensional matrices, the group operation is however not matrix multiplication. But in this case, the Lie group is still *isomorphic to a matrix group*, namely the multiplicative group $(\mathbb{R}^+, \cdot)$, with the isomorphism $\exp : \mathbb{R} \to \mathbb{R}^+$. Such Lie groups are called *linear*. For example, all finite groups are linear.
2. Some groups are 'too big' to be linear, for instance the group of permutations of $\mathbb{N}$.
3. There are also some connected Lie groups that are not linear, for instance the *metaplectic groups*, which are two-sheeted covers of the symplectic groups (see, e.g., Carter, Segal and Macdonald [CSM], page 130). These groups play a role in quantum mechanics. ◊

Generally, for a Lie algebra $\mathfrak{g}$ of a Lie group G, the exponential map is defined as follows. The elements $\xi \in \mathfrak{g}$ are understood as left invariant vector fields on G. As such they generate flows

$$\Phi^{(\xi)} : \mathbb{R} \times G \to G \qquad (\xi \in \mathfrak{g}),$$

which are complete because the group structure guarantees the existence of a time interval on which the Picard iteration converges that is independent of the initial condition $g \in G$. In terms of these flows, the exponential map is defined as

$$\exp : \mathfrak{g} \to G \quad , \quad \xi \mapsto \Phi^{(\xi)}(1, e)\,. \tag{E.3.1}$$

The exponential map maps small neighborhoods of $0 \in \mathfrak{g}$ diffeomorphically onto neighborhoods of $e \in G$. Namely for linear groups, one has $\mathrm{D}\exp(0) = \mathbb{1}$. However, it does not always map $\mathfrak{g}$ onto the connected component of $e \in G$.

This can be seen in the example of the group $\mathrm{SL}(2, \mathbb{R}) = \mathrm{Sp}(2, \mathbb{R})$, which was discussed in Exercise 6.26.d and illustrated on page 97. Among the hyperbolic matrices, the connected component of those matrices whose eigenvalues are all negative is not in the image $\exp\bigl(\mathfrak{sl}(2, \mathbb{R})\bigr)$ of the exponential map.

See also Abraham and Marsden [AM], *Example* 4.1.9, for the group $\mathrm{GL}(2, \mathbb{R})$.

E.25 Exercises (Exponential Map for GL(n, $\mathbb{R}$)) Show that the left resp. right invariant vector fields on the matrix group $\mathrm{GL}(n, \mathbb{R})$ are of the form

$$g \mapsto X^{(\xi)}(g) = g\,\xi \quad \text{resp.} \quad g \mapsto \xi\, g \qquad \bigl(g \in \mathrm{GL}(n, \mathbb{R}),\ \xi \in \mathrm{Mat}(n, \mathbb{R})\bigr)\,,$$

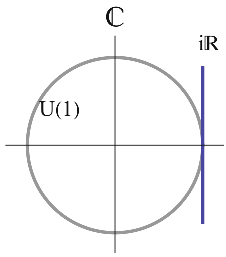
and that the commutator of two left invariant vector fields is given by

$$\left[X^{(\xi)}, X^{(\eta)}\right] = X^{([\xi,\eta])} \quad \text{with} \quad [\xi, \eta] = \xi\eta - \eta\xi \qquad \big(\xi, \eta \in \mathrm{Mat}(n, \mathbb{R})\big)\,,$$

and also that Definition (E.3.1) of $\exp(\xi)$ coincides with the matrix exponential (4.1). ◊

E.26 Example

1. The left and right invariant vector fields of an abelian Lie group are the same, and their Lie bracket vanishes. For instance on $\mathbb{R}^n$, it is exactly the constant vector fields that belong to the Lie algebra $\mathcal{X}_L(\mathbb{R}^n) = \mathcal{X}_R(\mathbb{R}^n)$.
   For the Lie group $U(1) = \{c \in \mathbb{C} \mid |c| = 1\}$ with its Lie algebra $\imath\,\mathbb{R}$, the figure on the right applies.



2. The matrix group $\mathrm{GL}(n, \mathbb{K})$ is open in $\mathrm{Mat}(n, \mathbb{K})$, its Lie algebra is therefore the tangent space $\mathfrak{gl}(n) = \mathrm{Mat}(n, \mathbb{K})$.
   The relation between a Lie algebra and the tangent space at the identity as described above permits a simple calculation of the matrix Lie algebra of a matrix Lie group, as can be seen in the example of $\mathrm{SO}(n, \mathbb{R}) \subset \mathrm{GL}(n, \mathbb{R})$: For a C^1 curve $A : (-\epsilon, \epsilon) \to \mathrm{SO}(n, \mathbb{R})$ with $A(0) = \mathbb{1}$, we have $A(s)A(s)^\top = \mathbb{1}$ and $\det\big(A(s)\big) = 1$ for all $s \in (-\epsilon, \epsilon)$, and therefore

$$0 = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} A(s)A(s)^\top = \dot{A}(0)A(0)^\top + A(0)\dot{A}(0)^\top = \dot{A}(0) + \dot{A}(0)^\top\,.$$

   In conclusion, introducing the vector space $\mathrm{Alt}(n, \mathbb{R}) := \{X \in \mathrm{Mat}(n, \mathbb{R}) \mid X^\top = -X\}$ of antisymmetric matrices, we have

$$\mathfrak{so}(n) = \mathrm{Alt}(n, \mathbb{R})\,, \qquad \text{(E.3.2)}$$

   because $X + X^\top = 0$ implies already that $\exp(X)\,\exp(X^\top) = \mathbb{1}$ and $\mathrm{tr}\,(X) = 0$, hence $\det(\exp X) = 1$. ◊

E.27 Exercises (Lie Groups and Lie Algebras)

1. In analogy to the above examples, calculate the Lie algebras $\mathfrak{u}(n)$, $\mathfrak{su}(n)$ and $\mathfrak{sp}(\mathbb{R}^{2n})$ of $\mathrm{U}(n)$, $\mathrm{SU}(n)$ and $\mathrm{Sp}(\mathbb{R}^{2n})$ respectively (see Example E.19.3 and (6.2.5)), and determine their dimensions.
2. *(Isomorphisms of Lie algebras)*
   (a) Show that the Lie algebra $\mathfrak{so}(3)$ with (13.4.8) is isomorphic to the Lie algebra $\mathbb{R}^3$, equipped with the vector product as the Lie bracket.
   (b) Show the isomorphism $\mathfrak{so}(3) \cong \mathfrak{su}(2)$, with

$$\mathfrak{su}(n) := \{X \in \mathrm{Mat}(n, \mathbb{C}) \mid X + X^* = 0,\ \mathrm{tr}(X) = 0\}\,.$$

(c) Denote by $\mathbb{H}$ the skew field of *quaternions* $q = a + bi + cj + dk$ with $a, b, c, d \in \mathbb{R}$, where i, j, k satisfy the following relations:

$$ij = k = -ji\ ,\ jk = i = -kj\ ,\ ki = j = -ik \text{ and } i^2 = j^2 = k^2 = -1\,,$$

and where conjugates are defined by $q^* := a - bi - cj - dk$ (see also KOECHER and REMMERT [KR]).
Show that $\mathfrak{su}(2)$ as a Lie algebra is isomorphic to the set $\mathrm{Im}\mathbb{H} := \{q \in \mathbb{H} \mid a = 0\}$ of *imaginary quaternions* with the commutator as the Lie bracket. To this end, use the imbedding

$$\mathbb{H} \to \mathrm{Mat}(2, \mathbb{C}) \quad , \quad q \mapsto \begin{pmatrix} a+b\imath & c+d\imath \\ -c+d\imath & a-b\imath \end{pmatrix}. \diamondsuit$$

E.28 Example (Parametrization of SO(3))
Denoting the ball as $B_r^d = \{x \in \mathbb{R}^d \mid \|x\| \le r\}$, the smooth mapping given by

$$A\colon B_\pi^3 \to \mathrm{SO}(3),\, x \mapsto \exp\bigl(i(x)\bigr)\ , \text{with } i : \mathbb{R}^3 \to \mathfrak{so}(3),\ \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \mapsto \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$$

is a parametrization of SO(3) (see also (13.4.8)). As x is in the kernel of the matrix $i(x)$, we infer that $A(x)$ is a rotation about the axis $\mathrm{span}(x)$. Since the eigenvalues are $\mathrm{spec}(i(x)) = \bigl\{0, \imath\|x\|, -\imath\|x\|\bigr\}$, the angle of rotation is $\|x\|$.

As every rotation in $\mathbb{R}^3$ can be viewed as a rotation to the right by an angle in the interval $[0, \pi]$, the mapping A is surjective.

Except for the identification of antipodes on the surface of the ball, namely $i(x) = i(-x)$ for $\|x\| = \pi$, the mapping is injective.

These properties of A also follow from the *Rodrigues formula*

$$\exp\bigl(i(x)\bigr) \;=\; 1\!\!1_3 \;+\; \tfrac{\sin(\|x\|)}{\|x\|}\; i(x) \;+\; \tfrac{1}{2}\left(\tfrac{\sin(\|x\|/2)}{\|x\|/2}\; i(x)\right)^2 \qquad \bigl(x \in \mathbb{R}^3 \setminus \{0\}\bigr). \tag{E.3.3}$$

This formula is proved by plugging

$$i(x)^2 = xx^\top - \|x\|^2 1\!\!1_3 \quad \text{, i.e.,} \quad i(x)^3 = -\|x\|^2\, i(x)$$

into the power series for exp and sorting for even and odd powers. $\diamondsuit$

This parametrization implies (see Exercise 6.53) that SO(3) is diffeomorphic to the real projective space $\mathbb{RP}(3) \cong S^3/\{\pm 1\!\!1\}$. On the other hand, by Example E.19.3, the Lie group SU(2) is diffeomorphic to the sphere S^3. This yields the two-sheeted covering of SO(3) by the group SU(2).

E.29 Theorem (SU(2) and SO(3)) *Let σ denote the linear isomorphism*

$$\sigma : \mathbb{R}^3 \to \mathfrak{su}(2) \quad , \quad \sigma(x) := \tfrac{1}{2}\begin{pmatrix} -\imath x_3 & -\imath x_1 - x_2 \\ -\imath x_1 + x_2 & \imath x_3 \end{pmatrix}.$$

Then the adjoint representation

$$\Pi : SU(2) \to SO(3) \quad , \quad \Pi_U(x) = \sigma^{-1}\big(U\sigma(x)U^{-1}\big) \qquad \big(U \in SU(2),\ x \in \mathbb{R}^3\big)$$

is a surjective group homomorphism with kernel $\{\pm 1\!\!1\}$.

Proof:

- Π is a group homomorphism with $\{\pm 1\!\!1\} \subseteq \ker(\Pi)$, because $\Pi_{-1\!\!1}(x) = \Pi_{1\!\!1}(x) = x$, and for $U, V \in \mathrm{SU}(2)$, one has

$$\Pi_U \circ \Pi_V(x) = \Pi_U\big(\sigma^{-1}\big(V\sigma(x)V^{-1}\big)\big) = \sigma^{-1}\big(UV\sigma(x)V^{-1}U^{-1}\big) = \Pi_{UV}(x).$$

  On the other hand, $U \in \ker(\Pi)$ if and only if $U\,v\,U^{-1} = v$ for all $v \in \mathfrak{su}(2)$. This is tantamount to $Uv = vU$ for all $v \in \mathfrak{su}(2)$, i.e., U is a multiple of unity.
- $\Pi_U \in \mathrm{GL}^+(3,\mathbb{R})$ since $\Pi_{1\!\!1} = 1\!\!1 \in \mathrm{GL}^+(3,\mathbb{R})$ and since $\mathrm{SU}(2) \cong S^3$ is connected. As $\mathrm{tr}\big(\sigma(x)\sigma(y)\big) = -\frac{1}{2}\langle x, y\rangle$, it follows: for that $\Pi_U \in \mathrm{SO}(3)$ already, because for $x, y \in \mathbb{R}^3$,

$$\langle \Pi_U(x), \Pi_U(y)\rangle = -2\mathrm{tr}\big(U\sigma(x)U^{-1}U\sigma(y)U^{-1}\big) = -2\mathrm{tr}\big(\sigma(x)\sigma(y)\big) = \langle x, y\rangle\,.$$

- $i : \mathbb{R}^3 \to \mathfrak{so}(3)\,,\ \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \mapsto \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$ is an isomorphism of the Lie algebras according to (13.4.8), and so is $\sigma : \mathbb{R}^3 \to \mathfrak{su}(2)$; therefore,

$$\psi := i \circ \sigma^{-1} : \mathfrak{su}(2) \to \mathfrak{so}(3)$$

  is an isomorphism, too (as was already shown in Exercise E.27.2).
  Thus the linearization $T_e\Pi : \mathfrak{su}(2) \to \mathfrak{so}(3)$ of the group homomorphism Π at the identity is of the form $T_e\Pi = \psi$, because, using that $v \times x = i(v)x$, it follows

$$T_e\Pi(u)(x) = \sigma^{-1}\big(u\sigma(x) - \sigma(x)u\big) = \sigma^{-1}(u) \times x = \psi(u)x \quad \big(u \in \mathfrak{su}(2),\ x \in \mathbb{R}^3\big).$$

- This shows that Π is a *local diffeomorphism*.[16] So the image $\Pi\big(\mathrm{SU}(2)\big) \subseteq \mathrm{SO}(3)$ is both open and closed. As the groups $\mathrm{SO}(n)$ are connected by Example E.19.2, we have $\Pi\big(\mathrm{SU}(2)\big) = \mathrm{SO}(3)$. □

E.4 Actions of Lie Groups

Just like any group can operate on a set (see Definition E.12), a topological group can operate on a topological space, or a Lie group on a manifold. But in these cases,

[16] In other words, according to Definition 2.36, this means that every point $x \in \mathrm{SU}(2)$ has a neighborhood U_x such that $\Pi\restriction_{U_x}$ is a diffeomorphism onto $\Pi(U_x)$.

one would desire compatibility of the structures. So in the first case, one will assume the continuity of the group action; in the second case, its continuous differentiability.

E.30 Example (Actions of Lie Groups)

1. Each Lie group G acts on itself from the right and from the left (E.1.3), and also by conjugation $I : G \times G \to G$ (E.1.4), i.e., by group automorphisms.
2. Consequently, it also acts on its Lie algebra $\mathfrak{g}$. Indeed, if we identify this Lie algebra with the tangential space T_eG at the identity $e \in G$, then with $I_g(e) = e$ $(g \in G)$, the *adjoint representation*

$$\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g} \quad , \quad \mathrm{Ad}_g(\xi) = \mathrm{D}I_g(e)\,\xi \qquad (g \in G,\ \xi \in T_eG \cong \mathfrak{g}) \tag{E.4.1}$$

is a linear group action of G on $\mathfrak{g}$. If the Lie group G is a matrix group, then

$$\mathrm{Ad}_g(\xi) = g\,\xi\,g^{-1} \qquad (g \in G,\ \xi \in T_eG).$$

3. This also defines a representation of G on the dual Lie algebra $\mathfrak{g}^*$. The dual mapping Ad_g^* of the linear mapping Ad_g is defined by the property

$$\big\langle \mathrm{Ad}_g^*(\xi^*), \eta\big\rangle = \big\langle \xi^*, \mathrm{Ad}_g(\eta)\big\rangle \qquad (\eta \in \mathfrak{g},\ \xi^* \in \mathfrak{g}^*).$$

But the mapping $g \mapsto \mathrm{Ad}_g^*$ is a right action, whereas we are using left actions. Accordingly, the *coadjoint representation* of G on $\mathfrak{g}^*$, defined by

$$\mathrm{Ad}^* : G \to \mathrm{Lin}(\mathfrak{g}^*) \quad , \quad g \mapsto \mathrm{Ad}^*_{g^{-1}}\,, \tag{E.4.2}$$

is a left action.
4. The differentiable dynamical systems, which we consider in this book, are also actions of Lie groups (namely of the Lie groups $\mathbb{Z}$ or $\mathbb{R}$, see Definition 2.43). ♢

E.31 Exercises (Adjoint Representation)
Show that for $G = \mathrm{SO}(3)$, and identifying $\mathfrak{so}(3)$ with $\mathbb{R}^3$ by (13.4.8), the adjoint representation of the rotation matrix $O \in \mathrm{SO}(3)$ is of the form $\mathrm{Ad}_O(\xi) = O\,\xi$. ♢

A special feature, which distinguishes actions of Lie groups $\Phi : G \times M \to M$ as compared to other group actions, comes along with the differentiability of the group action: it is the following relation between a Lie algebra $\mathfrak{g}$ and vector fields on the manifold M.

E.32 Definition
For an action of a Lie group $\Phi : G \times M \to M$ and $\xi \in \mathfrak{g}$, the vector field

$$X_\xi : M \to TM \quad , \quad X_\xi(m) := \frac{\mathrm{d}}{\mathrm{d}t}\Phi\big(\exp(t\xi), m\big)\Big|_{t=0}$$

on M is called the ***infinitesimal generator*** *of the group action generated by ξ.*

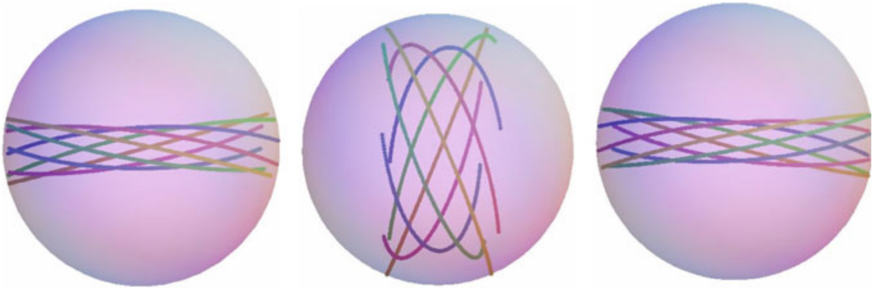


Figure E.4.1 Left and center: Orbits of the left action by a 1-parameter subgroup of SO(3), right: Right action by the same subgroup (note however the reversed torsion of the orbits). Shown in terms of the parametrization of $SO(3)$ by a ball as in Example E.28, page 557

For the left action $\Phi : G \times G \to G,\ \Phi_g(h) = g \circ h$, we obtain the *right* invariant vector fields on G. In Figure E.4.1, one can see the example of orbits for the left action generated by $\xi = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \in \mathfrak{so}(3)$; the vector field $X_\xi : \mathrm{SO}(3) \to T\mathrm{SO}(3)$ is tangential to these orbits.

It is no surprise that the adjoint representation is compatible with the Lie bracket and the exponential map:

E.33 Theorem *For every Lie group G with Lie algebra $\mathfrak{g}$, and for all $\xi, \eta \in \mathfrak{g}$,*

$$Ad_g\big([\xi, \eta]\big) = \big[Ad_g(\xi), Ad_g(\xi)\big]\ ,\ \exp\big(Ad_g(\xi)\big) = g \circ \exp(\xi) \circ g^{-1} \qquad (g \in G)$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t} Ad_{\exp(t\xi)}(\eta)\Big|_{t=0} = [\xi, \eta]\ . \tag{E.4.3}$$

E.34 Exercises (Adjoint Action)
Check these formulas for the Lie groups $\mathrm{GL}(n, \mathbb{R})$ and their corresponding Lie algebras $\mathrm{Mat}(n, \mathbb{R})$ (and thus for all matrix Lie groups $G \leq \mathrm{GL}(n, \mathbb{R})$). ◊

Identity (E.4.3) leads to the definition of the linear ad *operators*

$$\mathrm{ad}_\xi := \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ad}_{\exp(t\xi)}\Big|_{t=0} : \mathfrak{g} \to \mathfrak{g} \quad , \quad \eta \mapsto [\xi, \eta]\,. \tag{E.4.4}$$

They are the infinitesimal generators of the adjoint representation.

E.35 Definition

- *A continuous mapping $f : M \to N$ between topological spaces is called* ***proper*** *if the pre-images of compact sets are compact.*
- *A continuous group action $\Phi : G \times M \to M$ of a topological group G is called* ***proper*** *if the mapping $G \times M \to M \times M,\ (g, m) \mapsto \big(m, \Phi(g, m)\big)$ is proper.*

One frequently studies the space of orbits of a Lie group action, and sometimes this space is a manifold.

E.36 Theorem (Manifolds of Orbits) *If $\psi : G \times M \to M$ is a free and proper action of a Lie group G on a manifold M, then the quotient space*

$$B := M/G = \{\mathcal{O}(m) \mid m \in M\}$$

has the differentiable structure of a manifold, and the mapping

$$\pi : M \to B \quad , \quad x \mapsto \mathcal{O}(x) \, ,$$

which assigns to points of M their orbits, is a surjective submersion (see p. 501).

Proof: This is *Proposition 4.1.23* in ABRAHAM and MARSDEN [AM]. □

B can always be equipped with the quotient topology from Example A.2.4. But the following examples show that the requirement of a free and proper action cannot be omitted, if B is to be a manifold:

E.37 Exercises (Lie Group Actions)

1. Show that for $n \in \mathbb{N} \setminus \{1\}$, the mapping

$$\Phi : \mathrm{SO}(n) \times \mathbb{R}^n \to \mathbb{R}^n \quad , \quad (O, m) \mapsto Om$$

is a proper group action of the rotation group $\mathrm{SO}(n)$, but not a free one; and that $B = \mathbb{R}^n/\mathrm{SO}(n)$ is not a manifold (i.e., not a manifold without boundary).

2. Show that the mapping

$$\Phi : \mathbb{R} \times M \to M \quad , \quad (t, m) \mapsto \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} x$$

on $M := \mathbb{R}^2 \backslash \{0\}$ is a free, but not proper, action of the Lie group $(\mathbb{R}, +)$, and that B is not a Hausdorff space. ◊

Appendix F
Bundles, Connection, Curvature

F.1 Fiber Bundles

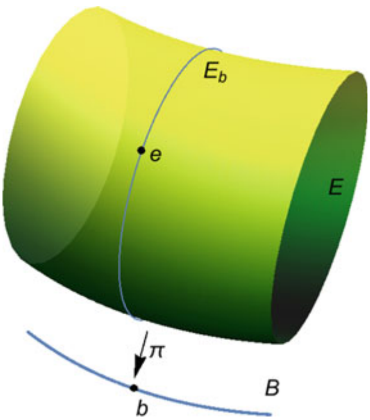
The cartesian product $E := B \times F$ of two manifolds is itself a manifold. If we denote by $\pi : E \to B\ ,\ (b, f) \mapsto b$ the projection onto the first factor, then (E, B, F, π) is an example of a fiber bundle.

F.1 Definition

- *Let E, B and F be topological Hausdorff spaces and*

$$\pi : E \to B$$

a continuous surjective mapping. Then (E, B, F, π) is called a (topological) ***fiber bundle*** *with* ***total space*** *E,* ***base space*** *B, and (****standard****)* ***fiber*** *F (see the figure to the right) if the* ***projection*** *π is* ***locally trivial****, which means that for all $b \in B$ there exists an open neighborhood $U \subseteq B$ and a homeomorphism $\Phi : \pi^{-1}(U) \to U \times F$ such that*



$$\Phi\big(\pi^{-1}(b')\big) = \{b'\} \times F \qquad (b' \in U).$$

- *For a fiber bundle (E, B, F, π), the family $(U_i, \Phi_i)_{i \in I}$ of such open sets $U_i \subseteq B$ and homeomorphisms Φ_i with $\bigcup_{i \in I} U_i = B$ is called a* ***local trivialization*** *of the fiber bundle.*
- *Analogously, for $r \in \mathbb{N}$ or $r = \infty$, one can talk about a C^r-****fiber bundle*** *if the total space E, the base space B, and the fiber F are C^r-manifolds and π as well as the $\Phi_i^{\pm 1}$ are C^r mappings.*
- *A continuous (or C^r respectively) mapping $S : B \to E$ is called a* ***section*** *if*

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

$$\pi \circ S = Id_B .$$

F.2 Remark

1. Instead of the somewhat clumsy naming (E, B, F, π), one frequently encounters the notation $\pi : E \to B$ or—*pars pro toto*—simply E. The *fibers* $\pi^{-1}(b) \cong F$ are also denoted as E_b.
2. In the introductory example of a product bundle, $(U, \Phi) := (B, \mathrm{Id}_E)$ is a local trivialization. Such bundles are called *trivial*. For example the parallelizable (see Definition A.43) tangent bundles are trivial. ♢

F.3 Example (One Dimensional Bundles)

1. A simple example of a nontrivial C^∞-fiber bundle is given by $E := S^1 \subset \mathbb{R}^2$, $B := \mathbb{R}\mathrm{P}(1)$ (the 1-dimensional real projective space consisting of all lines through the origin in $\mathbb{R}^2$, see Definition 6.50), and the mapping $\pi : E \to B$ that assigns to each point $x \in S^1$ on the circle the line through the origin that passes through that point.
   In this example, the standard fiber $F := \{-1, +1\}$ has two elements, but the total space E is connected, hence not homeomorphic to $B \times F$.
   An example of a local trivialization is $(U_i, \Phi_i)_{i \in I}$ with the index set $I := \{1, 2\}$ and
   $$U_i := \left\{\mathrm{span}(x) \mid x \in S^1 \subset \mathbb{R}^2,\ x_i > 0\right\},$$
   $$\Phi_i : \pi^{-1}(U_i) \to B \times F \quad , \quad x \mapsto \left(\mathrm{span}(x), \mathrm{sign}(x_i)\right).$$
   More generally, for $n \in \mathbb{N}$, we obtain a nontrivial fiber bundle with fiber $F = \{-1, +1\}$,
   $$\pi : S^n \to \mathbb{R}\mathrm{P}(n).$$
2. Another simple example of a C^∞-fiber bundle is the wrapping of the straight line onto the circle; the total space is $E := \mathbb{R}$, the base space is $B := S^1 \subset \mathbb{C}$, the projection is
   $$\pi : E \to B \quad , \quad x \mapsto \exp(2\pi \imath x),$$
   and the standard fiber is $F := \mathbb{Z}$. ♢

In cases like the above, where the standard fiber F is a discrete topological space, the fiber bundle is called a *covering*.

Principal Bundles and Vector Bundles

Fiber bundles frequently come with additional structures.

F.4 Definition *If the standard fiber F of the fiber bundle (E, B, F, π) is a topological group and there is a continuous (right) action*

$$\Psi : E \times F \to E \quad , \quad \Psi_f(e) := \Psi(e, f)$$

that leaves the fibers invariant (i.e., for all $f \in F$: $\pi \circ \Psi_f = \pi$) and is free and transitive on the fibers, then (E, B, F, π) is called a ***principal fiber bundle*** *or* ***principal bundle*** *with the* ***structure group*** *F.*

F.5 Remark By assumption, for any two points $e_1, e_2 \in E$ in the same fiber ($\pi(e_1) = \pi(e_2)$), there exists exactly one group element $f \in F$ such that $\Psi(e_1, f) = e_2$; therefore the space E/F of orbits for the group action is indeed homeomorphic to the base space B. If the topological spaces are differentiable manifolds, one assumes F to be a Lie group.

Frequently, principal bundles arise conversely from a free and proper action Ψ of a group G on a space E. Then Theorem E.36, specialized to manifolds, can be used to define the base manifold $B := E/G$. ◊

An example for such a group action is the action by $G = S^1$ on the energy surface $E := S^{2d-1} \subset \mathbb{R}^{2d}$ of a harmonic oscillator with d degrees of freedom, see Theorem 6.35.

Even though the fibers E_b of a principal bundle are homeomorphic to the topological group F for all points $b \in B$ in the base space, there is in general no fiberwise group multiplication that is continuous on E.

F.6 Example (Unit Tangent Bundle) For the n-sphere $B := S^n$, we will call

$$E := \{(x, y) \in S^n \times \mathbb{R}^{n+1} \mid \langle x, y\rangle = 0, \|y\| = 1\}$$

the *unit tangent bundle*, with the projection $\pi : E \to B,\ (x, y) \mapsto x$ and the fiber $F = S^{n-1}$. If especially $n = 2$, then F is a (Lie) group, and it acts by rotating the unit tangent vector y about the axis defined by x. If there existed a group multiplication in the fibers that were defined continuously across the fibers, it would distinguish one element in each fiber (the neutral element), which would give us a section $B \to E$. Such a section however does not exist; see Example A.44.2. ◊

F.7 Exercises (Triviality of Principal Bundles) Prove that a principal bundle (E, B, F, π) is trivial if and only if it has a section $B \to E$. ◊

Vector bundles are another type of fiber bundles. For them, the typical fiber F is a vector space. For example, in the case of the tangent bundle $E := TB$ of a manifold B, explained in Appendix A.3, the fibers $E_b = T_b B$ are real vector spaces of dimension $\dim(B)$.

F.8 Remark (Zero Section) In a vector bundle, the fibers are always (additive) groups, being vector spaces, but there is not in general a group action of F on E, because then by Exercise F.7, vector bundles would always be trivial due to the existence of the *zero section* $B \to E,\ b \mapsto 0 \in E_b$. But as in Example A.44.2 of the tangent bundle TS^2, vector bundles are not generally trivial. ◊

This is why the definition of vector bundles is not a special case of the definition of principal bundles.

F.9 Definition *A fiber bundle* (E, B, F, π) *in which the standard fiber* F *is a topological vector space is called a* ***vector bundle*** *(or* ***vector space bundle****) if the local trivializations* $(U_i, \Phi_i)_{i\in I}$ *for any two overlapping domains* $U_i, U_j \subseteq B$ *with* $U_{i,j} := U_i \cap U_j \neq \emptyset$ *can be related in terms of continuous* **transition functions** $t_{i,j} : U_{i,j} \to GL(F)$ *in the form*

$$\Phi_i \circ \Phi_j^{-1} : U_{i,j} \times F \longrightarrow U_{i,j} \times F \quad , \quad (b, f) \longmapsto \big(b,\ t_{i,j}(b) f\big) \ .$$

The ***rank*** *of the vector bundle is the dimension of its standard fiber* F.

In this book we only consider vector bundles whose standard fiber is a $\mathbb{K}$-vector space with $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

The fact that the transition between bundle charts is done by invertible linear mappings guarantees that the vector space structure of the fibers E_b is defined independently of the charts.

The usual operations of linear algebra can be carried out in vector bundles fiber by fiber, for example the factoring out of a subspace.

F.10 Example (Normal Bundle) Another class of vector bundles is given by *normal bundles*

$$T_M N / TM$$

of submanifolds $M \subseteq N$ (with $T_M N := \bigcup_{m\in M} T_m N$).
If N carries a Riemannian metric, this (*algebraic*) normal bundle is canonically isomorphic to the *geometric* normal bundle $TM^\perp$, which consists of all those tangent vectors of N that are locally orthogonal to M.

For instance if $M := S^n \subseteq \mathbb{R}^{n+1}$, one has

$$TM = \big\{(x, y) \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \mid \|x\| = 1,\ \langle y, x\rangle = 0\big\}\,,$$

and therefore

$$TM^\perp = \big\{(x, y) \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \mid \|x\| = 1,\ y = kx \text{ for some } k \in \mathbb{R}\big\}\,.\Diamond$$

F.11 Remark (Whitney Sum) If the operation of direct sum is applied fiberwise to two vector bundles $\pi^{(i)} : E^{(i)} \to B$ over the same base space B, so $(E^{(1)} \oplus E^{(2)})_b := E_b^{(1)} \oplus E_b^{(2)}$, one obtains the *direct sum*

$$\pi : E^{(1)} \oplus E^{(2)} \to B$$

of the vector bundles. This vector bundle is also called the *Whitney sum*.

So one has, for instance, $TM \oplus TM^\perp \cong T_M N$, where $TM^\perp$ is the normal bundle of a submanifold $M \subseteq N$ as in Example F.10. $\Diamond$

Orientation of Vector Bundles

The notion of *orientation* of an $n \in \mathbb{N}$-dimensional $\mathbb{R}$-vector space V carries over to vector bundles:

The set of bases $(e_1, \ldots, e_n)$ of V is parametrized by the general linear group $\mathrm{GL}(V)$ (see Example E.18), because for a second basis $(f_1, \ldots, f_n)$, there is exactly one $A \in \mathrm{GL}(V)$ for which $f_k = A(e_k)$ $(k = 1, \ldots, n)$, and conversely, the image of a basis under $A \in \mathrm{GL}(V)$ is again a basis.

This set of bases is partitioned into exactly two equivalence classes of bases that transform into each other by the subgroup $\mathrm{GL}^+(V)$. These two equivalence classes are called *orientations* of V.

As an example, for $\mathbb{R}^2$, the bases (e_1, e_2) and (e_2, e_1) are representatives of the two orientations; and for $\mathbb{R}^n$, the equivalence class of $(e_1, \ldots, e_n)$ is called the *standard orientation*.

In order to also cover the vector space $\mathbb{R}^0 = \{0\}$, one assigns to it the number 1 as *standard orientation*) or the number -1 as the other orientation.

F.12 Definition *A vector bundle* (E, B, F, π) *with a finite dimensional* $\mathbb{R}$*-vector space* F *as its standard fiber is called*

- ***oriented*** *if the fibers* F_b $(b \in B)$ *are given orientations that are constant in the local trivializations (see Definition F.1).*
- *The bundle is called* ***orientable*** *if it can be oriented in this sense.*
  A manifold M *is called* ***orientable*** *or* ***oriented*** *if its tangent bundle* TM *is* ***orientable*** *or* ***oriented*** *respectively.*

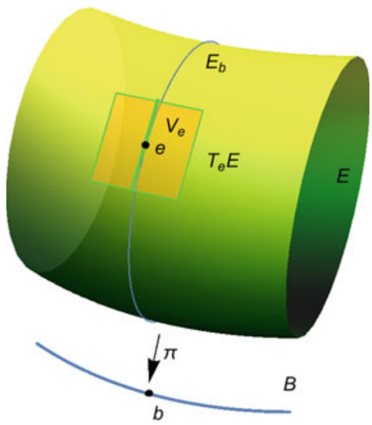
In this sense, the Möbius strip or the real projective spaces $\mathbb{R}\mathrm{P}(2k)$ of even dimension are not orientable; the $\mathbb{R}\mathrm{P}(2k+1)$ however are.

F.2 Connections on Fiber Bundles

In this section, we consider C^r-fiber bundles $\pi : E \to B$.
The linearization of the projection π is the mapping

$$T\pi : TE \to TB$$

from the tangent bundle of the total space E to the tangent bundle of the base manifold B.
For an arbitrary point $e \in E$ in the fiber E_b above $b := \pi(e)$, the induced mapping $T_eE \to T_bB$ of the tangent spaces is linear and surjective.
Its kernel $V_e \subset T_eE$ has the dimension $\dim(V_e) = \dim(F)$ of the standard fiber. Indeed, $V_e = T_eE_b$.
The notation V_e stands for the *vertical* subspace $T_eE_b \subset T_eE$, and $V \to E$ is called the *vertical bundle*, see the figure on the right.



F.13 Definition

- *An* ***(Ehresmann) connection*** *on the fiber bundle* (E, B, F, π) *is a smooth subbundle* H *of the tangent bundle* $TE \to E$ *such that the Whitney sum*

$$H \oplus V = TE\,.$$

- H *is called a* ***horizontal bundle****.*

So for every point $e \in E$, the horizontal subspace $H_e \subset T_eE$ complements the vertical subspace $V_e \subset T_eE$ in such a way that $H_e \oplus V_e = T_eE$, and the restriction of the linear mapping $T_e\pi : T_eE \to T_bB$ to H_e is an isomorphism.

F.14 Remark Subspaces of a vector space can be obtained as kernels (or as images) of linear mappings.

Applied to a fiber bundle (E, B, F, π), the vertical subspaces have already been described in this way. For an Ehresmann connection H on the fiber bundle, every tangent vector $X_e \in T_eE$ at the point $e \in E$ in the total space is decomposed uniquely into its *vertical* and *horizontal components* $\mathrm{ver}_e(X) \in V_e$, $\mathrm{hor}_e(X_e) \in H_e$:

$$X_e = \mathrm{ver}_e(X_e) \oplus \mathrm{hor}_e(X_e)\,. \tag{F.2.1}$$

Accordingly, a vector field $X : E \to TE$ has the components

$$\mathrm{ver}(X) : E \to V \quad \text{and} \quad \mathrm{hor}(X) : E \to H\,.$$

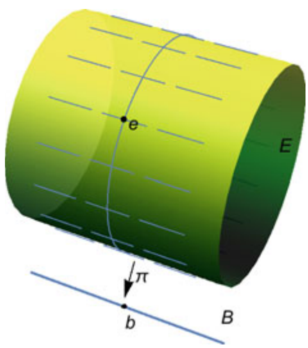
Conversely, a fiberwise linear mapping

$$\omega : TE \to V \quad , \quad T_eE \to V_e$$

defines an Ehresmann connection if and only if the linear mappings $\omega_e : T_eE \to V_e$ are projections onto the vertical space V_e, i.e., if

$$\omega_e \circ \omega_e = \omega_e \quad \text{and} \quad \omega_e(T_eE) = V_e \qquad (e \in E).$$ ♢

F.15 Example (Product Connection)



For a trivial bundle $\pi : E \to B$ with $E = B \times F$, the *product connection* is given by the kernel of $T\pi_2$, with the projection

$$\pi_2 : E \to F \quad , \quad (b, f) \mapsto f$$

on the standard fiber, see the figure on the right. ♢

In contrast to V however, a horizontal subbundle H is not defined by the bundle (E, B, F, π) itself, and accordingly there are many connections on a fiber bundle $\pi : E \to B$ (see Figure F.2.1).

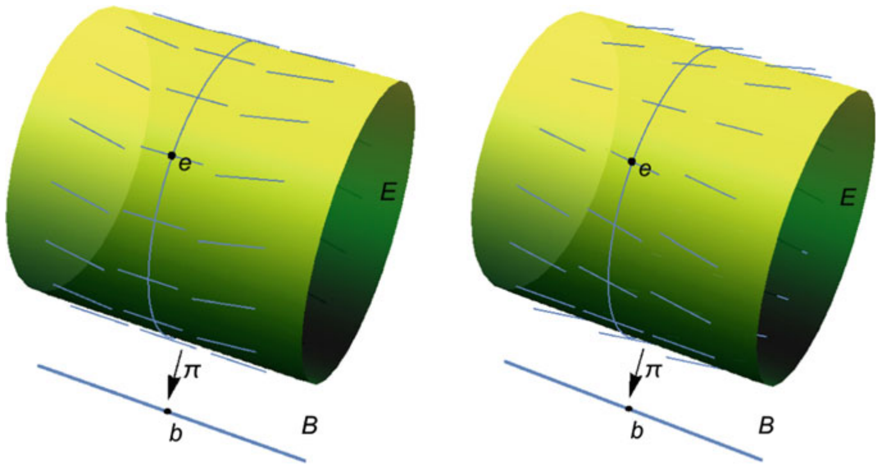


Figure F.2.1 Two different connections on the same bundle

F.16 Definition *Let $I \subseteq \mathbb{R}$ be an interval and $c \in C^1(I, B)$ a curve in the base space B of a C^1-fiber bundle $\pi : E \to B$.*

- *A curve $\tilde{c} \in C^1(I, E)$ is called* ***lift of*** *c if $c = \pi \circ \tilde{c}$.*
- *For a connection H on the fiber bundle, a lift $\tilde{c}$ of c is called* ***horizontal*** *if the velocity $\tilde{c}'$ is horizontal, i.e.,*

$$\tilde{c}'(t) \in H_{\tilde{c}(t)} \subset T_{\tilde{c}(t)} E \qquad (t \in I).$$

For each time $t \in I$ and each point $e \in E_{c(t)}$, there exists a unique vector $f_t(e) \in H_e$ that projects to $c'(t)$ under the linearized projection $T\pi$. The Picard-Lindelöf theorem (Theorem 3.17) now guarantees that for all $t_0 \in I$ and $e_0 \in E_{b(t_0)}$, the initial value problem

$$\tilde{c}'(t) = f_t\big(\tilde{c}(t)\big) \quad , \quad \tilde{c}(t_0) = e_0 \tag{F.2.2}$$

is uniquely solvable in an open neighborhood of t_0 in I. Frequently (for instance if the standard fiber is compact), there even exists a unique solution to (F.2.2) on all of I.

Connections on Principal Bundles and Vector Bundles

For principal bundles and for vector bundles, there exist special kinds of connections that are adjusted to the group, or to the vector space structure respectively.

We first study the case of principal bundles (Definition F.4), in which the Lie group F acts on the total space E as a group of fiber preserving diffeomorphisms Ψ_f.

F.17 Definition *An Ehresmann connection H on the principal bundle (E, B, F, π) is called an F-***connection*** if the linearized group action*

$$T\Psi_f : TE \to TE \qquad (f \in F)$$

satisfies

$$T_e\Psi_f(H_e) = H_{\Psi_f(e)} \qquad (e \in E,\ f \in F),$$

i.e., if the group action leaves the connection H invariant.

For example, the fiber bundle $\pi : E \to B$ in Figure F.2.1 can be interpreted as a principal bundle with the group $F = S^1$. Then only the connection in the left part of the figure is an F-connection.

As it is almost exclusively such F-connections that are used on principal bundles, the prefix 'F-' is frequently omitted.

If we denote by $\mathfrak{f}$ the Lie algebra of the Lie group F, then for all vectors $f \in \mathfrak{f}$, the vector field

$$X_f : E \to TE \quad , \quad X_f(e) = \frac{\mathrm{d}}{\mathrm{d}t}\Psi\big(e, \exp(tf)\big)\Big|_{t=0}$$

is vertical, i.e., $X_f(e) \in V_e$. Since the action of the Lie group F on a principal bundle is free and transitive, the linear mappings

$$\mathfrak{f} \to V_e \quad , \quad f \mapsto X_f(e) \qquad (e \in E)$$

are even bijective. So we can view the mapping $\omega : TE \to V$, which was introduced in Remark F.14 and represents the Ehresmann connection H, as an $\mathfrak{f}$-valued 1-form $\omega \in \Omega^1(E, \mathfrak{f})$ that satisfies

$$\omega(X_f) = f \qquad (f \in \mathfrak{f}). \tag{F.2.3}$$

F.18 Theorem *The Ehresmann connection defined by an $\omega \in \Omega^1(E, \mathfrak{f})$ satisfying (F.2.3) is an F-connection if*

$$ad_f\,\Psi_f^*\,\omega = \omega \qquad (f \in F) \tag{F.2.4}$$

with the adjoint representation ad of F in $\mathfrak{f}$.

Proof: See Kobayashi and Nomizu [KN], Chapter II, *Proposition 1.1.* □

F.19 Remark In the example of an abelian Lie group F (for example a torus), which is important for our applications, one has $\mathrm{ad}_f = \mathrm{Id}_{\mathfrak{f}}$, hence by (F.2.4), ω simply is invariant under the group action. ◊

Next, in order to introduce *linear* connections on vector bundles (E, B, F, π) (see Definition F.9), we use the following two smooth mappings.

– The multiplication by scalars $k \in \mathbb{K}$ on the fiber

$$M_k : E \to E \quad , \quad e \mapsto k\,e \qquad (k \in \mathbb{K}),$$

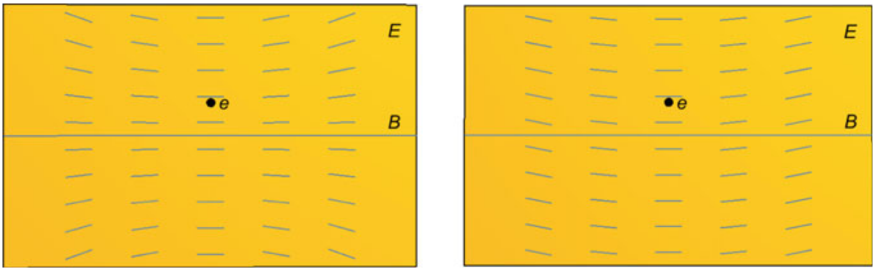


Figure F.2.2 A linear (left) and a non-linear (right) connection on a vector bundle $\pi : E \to B$

– the fiberwise vector addition on the Whitney sum $E \oplus E$:

$$A : E \oplus E \to E \quad , \quad (e_1, e_2) \mapsto e_1 + e_2 \, .$$

F.20 Definition *An (Ehresmann) connection H on a differentiable vector bundle (E, B, F, π) is called* ***linear****, if the horizontal subspaces transform as follows:*

1. $T_e M_k(H_e) = H_{ke} \qquad (e \in E,\ k \in \mathbb{K})$
2. $T_{(e_1,e_2)}A\ (H_{e_1}, H_{e_2}) = H_{e_1+e_2} \qquad \big((e_1, e_2) \in E \oplus E\big).$

A linear connection is shown in Figure F.2.2 (left), a (translation invariant) F-connection on the right. Most connections on vector bundles that occur in practice are linear.

F.21 Remark (Existence of Connections) For a differentiable fiber bundle (E, B, F, π), *there do exist* connections (and also F-connections respectively linear connections). Namely, the case of a trivial bundle with $E = B \times F$ and projection

$$\pi_2 : E \to F \quad , \quad (b, f) \mapsto f$$

onto the standard fiber, the *trivial connection* H satisfies all required properties.
In the general case, B has a partition of unity subordinate to the cover $(U_i)_{i \in I}$ (see Definition A.13), i.e., $\chi_i \in C(B, [0, 1])$ with $\mathrm{supp}(\chi_i) \subset U_i$ and $\sum_{i \in I} \chi_i = 1$. For an arbitrary point $e \in E$ of the total space with projection $b := \pi(e)$, we use the linearized projection

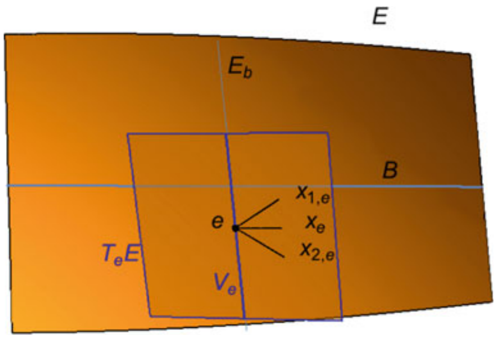
$$T_e\pi : T_eE \to T_bB \, .$$

If $y \in T_bB$, then for all indices $i \in I$ with $b \in U_i$, the trivial connection H_i has the property that the horizontal space $H_{i,e} \subset T_eE$ has exactly one point $x_i \in H_{i,e}$ that lies above y with respect to $T_e\pi$.

Likewise, the (finite!) convex combination

$$x := \sum_{i\in I} \chi_i(b)x_i$$

satisfies $T_e\pi(x) = y$. This way, we can take a convex combination of all the horizontal spaces $H_{i,e}$ and obtain the horizontal subspace $H_e \subset T_eE$, see the figure to the right. ♢



Convex combination of connections

In Section 8.4, the Christoffel symbols $\Gamma^h_{i,j}$ (with indices $i, j, h = 1, \ldots, d$) were introduced as

$$\Gamma^h_{i,j}(x) = \tfrac{1}{2}g^{h,k}(x)\left(\frac{\partial g_{k,j}}{\partial x_i}(x) + \frac{\partial g_{i,k}}{\partial x_j}(x) - \frac{\partial g_{i,j}}{\partial x_k}(x)\right) \ (x \in U).$$

In a chart with domain $U \subseteq M$, they are the coefficients of the geodesic equation (8.4.3) on the Riemannian manifold (M, g). This means that they define a connection on the tangent bundle $\pi : TM \to M$, called the Levi-Civita connection. Geodesic motion is then obtained by parallel translation of the velocity vector.

If $\partial_{\varphi_1}, \ldots, \partial_{\varphi_d}$ denote the coordinate vector fields in a chart (U, φ) of M, as introduced on page 523, then the vector fields $X \in \mathcal{X}(M)$ are locally of the form $X = \sum_{k=1}^d X_k\partial_{\varphi_k}$ with coefficient functions $X_k : U \to \mathbb{R}$.

The *covariant derivative*, i.e., the change of a vector field Y in the direction of a vector field X, can then be written in the following form (using Einstein's summation convention):

$$\nabla_X Y = \left(X_i\, \partial_{\varphi_i} Y_k \ + \ \Gamma^k_{i,j} X_i\, Y_j\right) \partial_{\varphi_k}\, .$$

F.22 Theorem (Principal Theorem of Riemannian Geometry)
The ***Levi-Civita connection*** *defined locally in this way is characterized by the following two properties:*

- *It is compatible with the metric g, i.e.,*

$$L_X\big(g(Y, Z)\big) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \qquad \big(X, Y, Z \in \mathcal{X}(M)\big);$$

- *It is torsion free, i.e.,* $\nabla_X Y - \nabla_Y X = [X, Y] \ \ \big(X, Y \in \mathcal{X}(M)\big)$.

F.3 Distributions and the Frobenius Theorem

In order to introduce the curvature of a connection, we first study properties of subbundles of the tangent bundle.

F.23 Definition • *A* ***(geometric) distribution*** *in a manifold M is a smooth subbundle $D \subseteq TM$ of the tangent bundle.*
rank(D) *is the rank of D as a vector bundle, i.e., the (constant) dimension of the fibers $D_x \subseteq T_xM \quad (x \in M)$.*

- *A* rank(D)*-dimensional submanifold $N \subseteq M$ is called an* ***integral manifold*** *of D if $T_xN = D_x \quad (x \in N)$.*
- *A distribution $D \subseteq TM$ of rank k is called* ***integrable*** *if every point $x \in M$ lies in some integral manifold of D (Figure F.3.1).*
- *It is called* ***involutive*** *if the commutator $[X, Y]$ of two vector fields $X, Y \in \mathcal{X}(M)$ that are tangential to D is again tangential to D.*

A k-dimensional subspace of an m-dimensional vector space can be described as the common zero set of $m - k$ linearly independent linear forms. Similarly, every distribution of rank k can locally (i.e., in appropriate neighborhoods $U \subset M$ of the $x \in M$) be described as the intersection of the kernels of independent 1-forms $\omega_1, \ldots, \omega_{m-k} \in \Omega^1(U)$.

F.24 Remark (Existence of Distributions)

- There may not exist distributions of rank k on a given manifold M.
An example is the sphere $M = S^2$ and $k = 1$. This is proved similarly as the hairy ball theorem (Example A.44.2).
- Even if there is a distribution of rank k on a manifold M, it is not necessarily representable *globally* as the intersection of kernels of independent 1-forms. An example is $M = S^2$ and $k = 0$, which is of course a rather trivial distribution. As was just noticed, there does not exist a nowhere vanishing 1-form on S^2. ◊

F.25 Theorem (Frobenius) *Let $E \subseteq TM$ be a geometric distribution of rank k on the m-dimensional manifold M. Then the following conditions are equivalent:*

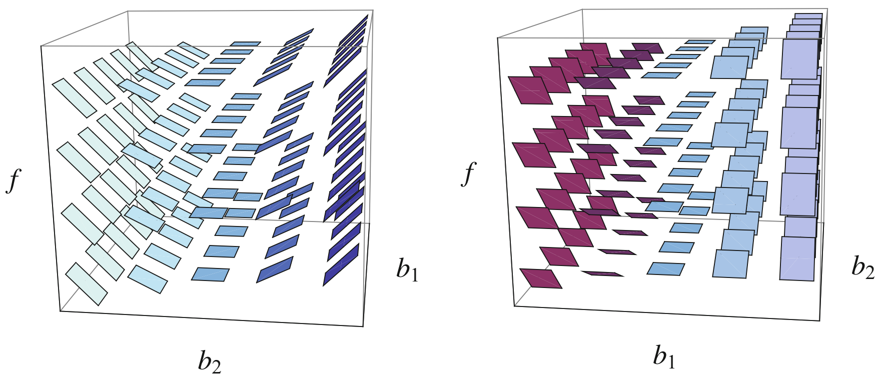


Figure F.3.1 Integrable (left) and non-integrable (right) distribution on the bundle $\pi : \mathbb{R}^3 \to \mathbb{R}^2$

1. *E is integrable;*
2. *E is involutive;*
3. *There exists a cover of M by neighborhoods $U \subseteq M$ such that for a local representation $E \cap TU = \{v \in TU \mid \omega_1(v) = \ldots = \omega_{m-k}(v) = 0\}$ of the distribution by $\omega_1, \ldots, \omega_{m-k} \in \Omega^1(U)$, there exist further 1-forms $\theta_{i,j} \in \Omega^1(U)$ with*

$$d\omega_i = \sum_{j=1}^{m-k} \theta_{i,j} \wedge \omega_j \qquad (i = 1, \ldots, m-k).$$

4. *In the notation of item 3, $d\omega_i \wedge \omega_1 \wedge \ldots \wedge \omega_{m-k} = 0$ $(i = 1, \ldots, m-k)$.*

The proof can be found in Chapter 4.1 of the book [AF] by AGRICOLA and FRIEDRICH. □

F.26 Example (Integrability of Geometric Distributions)

1. Distributions $D \subseteq TM$ of *rank one* are integrable. This follows from the principal theorem in the theory of differential equations (Theorem 3.45), because in an appropriate neighborhood $U \subseteq M$ of every point, we can find a non-vanishing smooth vector field $X \in \mathcal{X}(U)$ that is tangential to the distribution. Its orbits are integral manifolds of D.
2. The *contact distributions* from Remark 10.11 are not integrable. This is because a contact form $\omega \in \Omega^1(M)$ on a $(2n+1)$-dimensional manifold M generates a volume form $\omega \wedge (d\omega)^{\wedge n}$ by definition. But this is in contradiction to condition 3 in the Frobenius theorem ($d\omega = \theta \wedge \omega$), which implies $\omega \wedge d\omega = 0$ by antisymmetry.
3. An (*Ehresmann*) *connection* on a fiber bundle (Definition F.13) is a special case of a distribution. Its (possible) failure to be integrable leads to the notion of curvature, and thus into the center of differential geometry. ♢

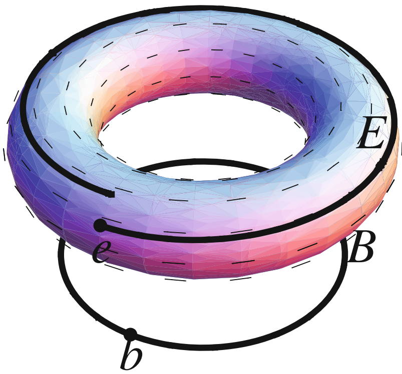
F.4 Holonomy and Curvature

We will assume for simplicity that in the bundle (E, B, F, π) under consideration with connection H, it is possible for arbitrary curves $c \in C^1(I, B)$ in the base space B to be lifted horizontally[17] (the same is then true for *piecewise* continuously differentiable curves).

[17] Sometimes this property, which is always satisfied when F is compact, is added as part of the definition of a connection. See for instance Chapter 9.9 of KOLÁR, MICHOR, and SLOVÁK [KMS].

Let I be the interval $[0, 1]$ and c closed ($c(1) = c(0) = b$). If the curve $\tilde{c}_e : I \to E$ denotes the horizontal lift of c with initial point $\tilde{c}_e(0) = e \in E_b$, then $\tilde{c}_e(1) \in E_b$ need not be equal to e (see the figure to the right). But at least, the fiber E_b above the initial point b of c is mapped onto itself homeomorphically under the mapping

$$\hat{c} : E_b \to E_b \quad , \quad e \mapsto \tilde{c}_e(1) .$$



Holonomy of a connection

F.27 Definition *For $b \in B$, we call*

$$Hol(b) := \left\{ \hat{c} : E_b \to E_b \mid c \in C^1(I, B), c(0) = c(1) = b \right\}$$

the ***holonomy group*** *of b.*

F.28 Remark

1. Hol(b) is indeed a group, because for two paths $c_1, c_2 : I \to B$, we can lift horizontally the composite path $c_1 * c_2 : I \to B$ defined by

$$c_1 * c_2(t) := \begin{cases} c_2(2t) & , t \in \left[0, \frac{1}{2}\right] \\ c_1(2t-1) & , t \in \left[\frac{1}{2}, 1\right] \end{cases} \tag{F.4.1}$$

   which corresponds to the composition of $\hat{c}_1$ and $\hat{c}_2$.[18] Likewise, the existence of the inverse of $\hat{c}$ is guaranteed by $(\hat{c})^{-1} = \widehat{c^{-1}}$ where $c^{-1}(t) := c(1-t)$.
2. If the base manifold B is connected, then all the groups Hol(b) ($b \in B$) are isomorphic to each other, because for $b_0, b_1 \in B$ there is a curve $d \in C^1(I, B)$ with $d(0) = b_0, d(1) = b_1$. If $c_1 \in C^1(I, B)$ starts and ends at b_1, then the curve $c_0 := d^{-1} * c_1 * d$ (defined analogously as in (F.4.1)) is a loop based at b_0. Then the mapping $\text{Hol}(b_1) \to \text{Hol}(b_0)$, $\hat{c}_1 \mapsto \hat{c}_0$ is a group isomorphism. ◊

F.29 Example (Three-Axes Stabilization)
Without the use of control nozzles, spacecraft cannot change their angular momentum. Normally, this angular momentum is very small so that the orientation in space stays approximately constant. To change this orientation, one employs reaction wheels, in the case of three-axes stabilization wheels with three orthogonal axes (see also the box on page 379).

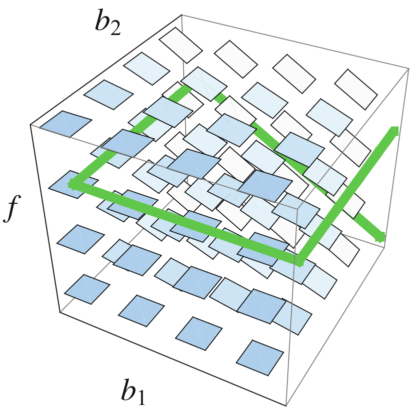
If the angle $\theta \in B := S^1$ gives the present angle of such a wheel relative to the satellite, then the orientations of satellite and wheel in space represent a point

[18] While in general, $c_1 * c_2$ is only piecewise continuously differentiable, we can achieve by reparametrization that $c_2'(1) = c_1'(0) = 0$. This reparametrization changes neither $\hat{c}_1$ nor $\hat{c}_2$.

$(\varphi_S, \varphi_R) \in E := \mathbb{T}^2$ on a torus. We view E as the total space of a bundle $\pi : E \to B$, with projection $\theta = \pi(\varphi_S, \varphi_R) := \varphi_R - \varphi_S$.

If the moments of inertia of satellite and wheel about the axis of the wheel are $I_S, I_R > 0$ (see page 314), then the total angular momentum about this axis is $I_S\dot{\varphi}_S + I_R\dot{\varphi}_R$. Thus conservation of the angular momentum represents geometrically a connection on E, namely the one defined by the 1-form $I_S\, d\varphi_S + I_R\, d\varphi_R = (I_S + I_R)\, d\varphi_S + I_R\, d\theta$ and depicted on page 574.

So the holonomy for the turn of the wheel by $\Delta\theta = 2\pi$ is, measured by the change in orientation of the satellite, $\Delta\varphi_S = -2\pi I_R/(I_S + I_R)$. ◊



As shown in this example, an integrable connection may have a nontrivial holonomy group if the base manifold is not simply connected.

There is however also the possibility that the holonomy group is nontrivial because the connection is not integrable (see the figure). Then the curvature of the connection is a quantitative measure for its lack of integrability.

The image suggests that in the limit of short curve length, the holonomy is proportional to the area enclosed by the projected curve in the base manifolds. The following definition makes this observation more precise and denotes the factor of proportionality as curvature.

F.30 Definition (Curvature of a Connection)
The curvature of the Ehresmann connection H is (with respect to the decomposition (F.2.1) of a vector field) the 2-form

$$K \in \Omega^2(E, TE) \quad , \quad K(X, Y) = \mathrm{ver}\big([\mathrm{hor}(X), \mathrm{hor}(Y)]\big) \qquad \big(X, Y \in \mathcal{X}(E)\big).$$

F.31 Remark (The Curvature as a Vector Valued 2-Form)

1. For a point $e \in E$, the tangent vector $K(X, Y)(e) \in T_e E$ thus defined depends only on $X(e)$ and $Y(e)$, so K is indeed a vector-valued 2-form.

   This is due to the fact that generally on a manifold E, Lie brackets of vector fields $U, V \in \mathcal{X}(E)$ satisfy the following relation with respect to multiplication by a function $f \in C^\infty(E, \mathbb{R})$:

$$[U, fV] = f[U, V] + df(U)\, V \,. \tag{F.4.2}$$

This is an immediate consequence of Definition 10.20.

The first term in the sum (F.4.2) does not involve a derivative of f at all. The second term does depend on df, but if V is horizontal, then so is its product with the function $df(U)$. Thus the second term vanishes in the projection to the vertical subspace.

The analogous statement applies for multiplication of U with a function.

2. The curvature of an F-connection on a principal bundle with an abelian Lie group is invariant under fiber translations (see Remark F.19). So it can also be viewed as a vector valued 2-form on the *base* manifold.
   This is the case in electrodynamics. There, space-time $\mathbb{R}^4$ is the base manifold, and the fiber is the abelian group $\mathrm{U}(1)$. The curvature consists of electric and magnetic field strength (see Example B.21 and THIRRING [Th2]). ◊

Appendix G
Morse Theory

Morse Theory connects the topology of an n-dimensional manifold M with the critical points of a function $f \in C^2(M, \mathbb{R})$.

G.1 Definition

- *A critical point* $x \in M$ *of* $f \in C^2(M, \mathbb{R})$ *is called* ***nondegenerate*** *if in an arbitrary chart at* x*, the Hessian matrix* $\mathrm{D}^2 f(x) \in \mathrm{Mat}(n, \mathbb{R})$ *is regular.*
- *The* ***index*** *Ind*(x) *of the critical point* $x \in M$ *of* f *is defined to be the index*[19] *of* $\mathrm{D}^2 f(x)$.
- $f \in C^2(M, \mathbb{R})$ *is called a* ***Morse function*** *if all its critical points are nondegenerate.*

The set $\mathrm{Crit}(f) \subseteq M$ of critical points of f is closed. If f is a Morse function, then this *critical set* is discrete. We begin by considering only compact manifolds M. On them, a Morse function has only finitely many critical points.

Since the subset of nondegenerate symmetric matrices is open and dense in the vector space $\mathrm{Sym}(n, \mathbb{R})$, the property of being a Morse function is generic (see Remark 2.44.2, and Theorem 1.2 in Chapter 6 of [Hirs]).

For example, if $M \subset \mathbb{R}^n$ is a submanifold, then for Lebesgue-almost all $a \in \mathbb{R}^n$, the restriction of the linear form $f_a : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \langle x, a\rangle$ to M is a Morse function (see *Proposition 17.18* in Bott and Tu [BT]).

G.1 Morse Inequalities

The *Morse inequalities* for a Morse function $f : M \to \mathbb{R}$ establish a relation between the topology of M (as encoded in its Betti numbers—see Equation B.8.2) and the

[19]The index of an $n \times n$ matrix was introduced in Definition 5.2 as the sum of the algebraic multiplicities of those eigenvalues λ that satisfy $\Re(\lambda) < 0$.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

cardinalities $\mathrm{crit}_\ell(f) := |\mathrm{Crit}_\ell(f)|$ of the sets $\mathrm{Crit}_\ell(f)$ of critical points with index ℓ. In their simplest form, they state for a compact manifold M^n that

$$\boxed{\mathrm{crit}_\ell(f) \geq \mathrm{betti}_\ell(M) \qquad (\ell = 0, \ldots, n).} \tag{G.1.1}$$

A Morse function $f : M \to \mathbb{R}$ that satisfies (G.1.1) with equality is called *perfect*.[20]

It is however important that for dimensions $n \geq 1$, we can make the left sides of the inequalities arbitrary large by changing f, but we cannot do it in any arbitrary manner. In particular, the following equation will always be satisfied:

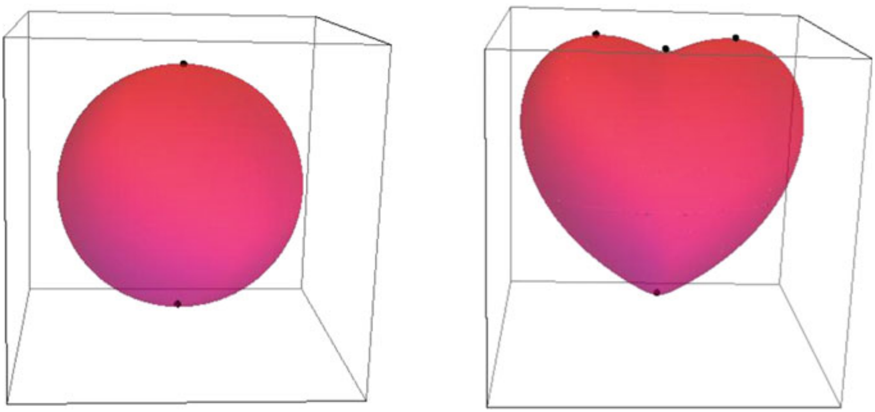
$$\boxed{\textstyle\sum_{\ell=0}^{n} (-1)^\ell \, \mathrm{crit}_\ell(f) = \sum_{\ell=0}^{n} (-1)^\ell \, \mathrm{betti}_\ell(M) =: \chi(M)\,,} \tag{G.1.2}$$

where we have used the alternating sum of the Betti numbers as our definition of the *Euler characteristic* $\chi(M)$ of the manifold.

G.2 Example (Spheres) $M = S^n$. The height function $f(x) := x_{n+1}$ of S^n with respect to the usual imbedding into $\mathbb{R}^{n+1}$ has one minimum and one maximum and no other critical points. Therefore $\mathrm{crit}_0(f) = \mathrm{crit}_n(f) = 1$, $\mathrm{crit}_1(f) = \ldots = \mathrm{crit}_{n-1}(f) = 0$ and hence

$$\chi(S^n) = 1 + (-1)^n \qquad (n \in \mathbb{N}_0).$$

If we make a (smooth) dent into a round sphere, another maximum may appear for the height function, but then it is unavoidable that another critical point appears as well.



The figure on the right shows for $n = 2$ the case $\mathrm{crit}_0(f) = \mathrm{crit}_{n-1}(f) = 1$, $\mathrm{crit}_n(f) = 2$. The alternating sum is not changed. ◊

[20] Such a function need not exist for M. An example is given by a 3-manifold called the Poincaré sphere, which is of significance in connection with the Poincaré conjecture; see *Remark* 5.15 in the book [Ni] by NICOLAESCU.

G.3 Remark

1. The Morse inequalities are used to conclude from critical points of appropriate Morse functions to the topology of the manifold. For instance, both the orientable and the non-orientable surfaces can be classified topologically by their Euler characteristics.

   It is of at least equal importance to show that Morse functions must have many critical points. One example is the question of the minimum number of periodic orbits for Hamiltonian systems, see the Arnol'd conjecture (page 480).
2. But it is not absolutely necessary for the manifold M to have finite dimension. For example, if I denotes an interval, the space $H^1(I, N)$ of H^1-curves[21] $c : I \to N$ on a Riemannian manifold (N, g) is not compact. But at least it has the structure of an infinite-dimensional manifold (see [Kli2], Chapter 2.3), with a Riemannian metric. For smooth vector fields v, w along c (i.e., $v(t), w(t) \in T_{c(t)}N$), one defines

$$\langle v, w\rangle := \int_I \Big[g_{c(t)}\big(v(t),\ w(t)\big) + g_{c(t)}\big(\nabla v(t),\ \nabla w(t)\big)\Big]\,\mathrm{d}t\,. \qquad \text{(G.1.3)}$$

   This gives $H^1(I, N)$ a structure called Hilbert manifold. The *energy functional*

$$\mathcal{E} : H^1(I, N) \to [0, \infty) \quad , \quad c \mapsto \tfrac{1}{2}\int_I g_{c(t)}\big(\dot c(t),\ \dot c(t)\big)\,\mathrm{d}t \qquad \text{(G.1.4)}$$

   is smooth. Its critical points are the constant curves, because the derivative

$$\mathrm{D}\mathcal{E}(c)(v) = \int_I g_{c(t)}\big(\dot c(t), \nabla v(t)\big)\,\mathrm{d}t \qquad \big(v \in T_c H^1(I, M)\big) \qquad \text{(G.1.5)}$$

   in direction $v := \dot c$ is otherwise positive. So the critical set is homeomorphic to N. ◊

There exist various proofs of the Morse inequalities (see also Remark 17.15.4).

G.4 Remark (Witten's Proof of the Morse Inequalities)
In 1982, in [Wit], the physicist Edward Witten published a proof of the Morse inequalities for closed orientable manifolds; his proof relies on a deformation of an operator known as the *Laplace-Beltrami operator* on the vector space $\Omega^*(M)$ by means of a Morse function. ($\Omega^*(M)$ was introduced in (B.4.1), see also [CFKS], Chapter 11.)

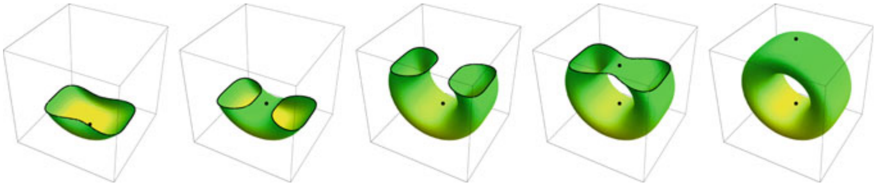
This proof is the starting point for interesting developments in mathematics and in quantum field theory. ◊

We gather briefly the main steps of the classical proof. In this proof, the manifold M is constructed successively from the *sublevel sets*

[21] which means those with square integrable first derivative.

$$M_b := f^{-1}\big((-\infty, b]\big) \qquad (b \in \mathbb{R})$$

of the Morse function by increasing the parameter b.



Sublevel sets $\mathbb{T}^2_b$ of a perfect Morse function $f : \mathbb{T}^2 \to \mathbb{R}$ on the torus.

In doing so, parameter intervals are distinguished depending on whether they do or do not contain a critical value.

G.5 Lemma *Let $f \in C^{r+1}(M, \mathbb{R})$, $r \geq 1$. If $M_{a,b} := f^{-1}\big([a, b]\big)$ does not contain a critical point, then the manifold with boundary $M_{a,b}$ is diffeomorphic to $f^{-1}(a) \times [a, b]$ (with a C^r-diffeomorphism that maps the level sets $f^{-1}(c)$ onto $f^{-1}(a) \times \{c\}$).*

Proof: On $M_{a,b}$, one uses the flow of the normalized gradient vector field $X := \frac{\nabla f}{\|\nabla f\|^2}$ with respect to an arbitrarily chosen Riemannian metric. Since $df(X) = 1$, this flow maps level sets into level sets. Details can be found in Chapter 6.1, Theorem 2.2. of Hirsch [Hirs]. □

G.6 Lemma **(Morse Lemma)** *Let $f \in C^{r+1}(M, \mathbb{R})$, $r \geq 1$, and let $m \in M^n$ be a nondegenerate critical point of f with index k.*

Then there exists a C^r-chart (U, φ) at m for which

$$f \circ \varphi^{-1}(x) = f(m) - \sum_{i=1}^{k} x_i^2 + \sum_{i=k+1}^{n} x_i^2 \qquad \big(x \in \varphi(U)\big).$$

Proof: In Hirsch [Hirs], Theorem 1.1 of Chapter 6.1 (Differentiability in Exercise 1 there). □

If $M_{a,b}$ contains exactly one critical point m, then topologically, $M_b = M_a \cup M_{a,b}$ arises[22] from M_a by attaching the stable manifold

$$W^s(m) := \big\{x \in M \mid \lim_{t \to +\infty} \Phi_t(x) = m\big\}$$

of m with respect to the gradient flow Φ (see Figure G.1.1, and §3 of Milnor [Mil]). This stable manifold has dimension $\mathrm{Ind}(m)$ by the Morse lemma.

By means of the techniques of singular homology, which are explained in the following section, one controls the change in topology that the sublevel set undergoes in this process.

---

[22] Namely M_a with the stable manifold attached is a deformation retract of M_b:
For a subset $A \subseteq B$ of a topological space B, a homotopy $g : B \times [0, 1] \to B$ is called a *deformation retraction* and A a *deformation retract* of B, if $g(B, 1) = A$ and $g(\cdot, 1)\restriction_A = \mathrm{Id}_A$. This implies in particular that A and B are homotopy equivalent.

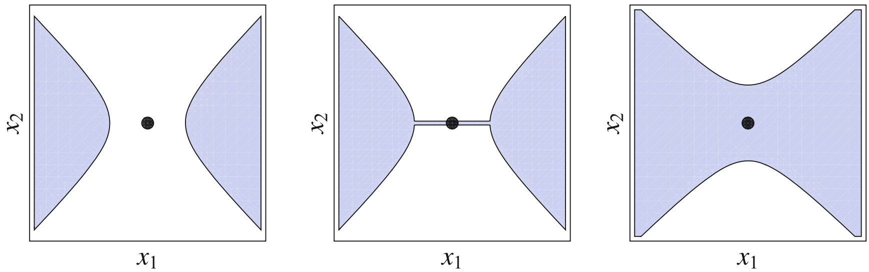


Figure G.1.1 Sublevel sets $M_{c-\varepsilon}$ (left) and $M_{c+\varepsilon}$ (right) of a Morse function $f : M \to \mathbb{R}$ near a critical point m with $c := f(m)$ and $\mathrm{Ind}(m) = 1$. The figure in the center shows the union of $M_{c-\varepsilon}$ and the stable manifold of m. The chart corresponds to the one in the Morse lemma

G.2 Singular Homology

The Morse inequalities (G.1.1) contain the Betti numbers, which we had introduced in (B.8.2) by means of de-Rham cohomology. It seems therefore natural to prove these inequalities by means of a calculus of differential forms.

But instead of de-Rham cohomology, the proof usually uses a similar tool called singular homology. A reason for this choice is the fact that singular homology can be defined for arbitrary topological spaces, not just for finite dimensional manifolds.

This ensures that Morse theory can also be applied to questions as important as the question of closed geodesics (see Remark G.3.2).

Singular homology is a homology theory that assigns to a topological space X abelian groups $H_k(X)$, $k \in \mathbb{N}_0$ that partly characterize the space. The book [Cr] by Croom is an elementary introduction to the area.

The building blocks are the simplices: If $a_0, \ldots, a_k \in \mathbb{R}^n$ are geometrically independent in the sense that they do not lie in any $(k-1)$-dimensional affine subspace, then their convex hull

$$\sigma_k := [a_0, \ldots, a_k] := \left\{ \sum_{i=0}^k t_i a_i \;\middle|\; t_i \geq 0,\ \sum_{i=0}^k t_i = 1 \right\} \subset \mathbb{R}^n$$

is called a **standard k-simplex**; the i^{th} **face**

$$\sigma_{k-1}^{(i)} := [a_0, \ldots, \hat{a}_i, \ldots, a_k] \qquad (i = 0, \ldots, k)$$

of σ_k is the convex hull of the vertices of σ_k other than a_i.

G.7 Definition

- *A* ***singular*** ***k-simplex*** *of the topological space X is a pair (σ_k, f) where f is a continuous mapping $f : \sigma_k \to X$.*

- *Let G be an abelian group. A* ***singular k-chain*** *is a finite formal linear combination* $\sum_i g_i\,(\sigma_{k,i}, f_i)$ *of singular k-simplices, with* $g_i \in G$*. The set* $C_k(X; G)$ *(or briefly:* $C_k(X)$*) of these singular chains is therefore again an abelian group.*
- *The* ***boundary*** *of a singular k-simplex* (σ_k, f) *is the formal linear combination*

$$\partial(\sigma_k, f) \;:=\; \sum_{q=0}^{k}(-1)^q \left(\sigma_{k-1}^{(q)}, f\restriction_{\sigma_{k-1}^{(q)}}\right) \;\in\; C_{k-1}(X; G)\,,$$

the ***boundary*** *of a singular k-chain* $c_k := \sum_i g_i\,(\sigma_{k,i}, f_i)$ *is*

$$\partial c_k := \sum_i g_i\,\partial(\sigma_{k,i}, f_i)\,.$$

G.8 Theorem $\partial : C_k(X) \to C_{k-1}(X)$ $(k \in \mathbb{N})$ *is a group homomorphism, and*

$$\boxed{\partial\partial = 0\,.}$$

This formula, which can be checked easily, is comparable to the property $dd = 0$ of the exterior derivative. Just as the latter was used to define de-Rham cohomology in (B.8.1), Theorem G.8 leads to the definition of singular homology:

G.9 Definition (and Theorem)

- *A chain* $c_k \in C_k(X; G)$ *is called a*
  - ***cycle*** *(*$c_k \in Z_k(X; G)$*) if* $\partial c_k = 0$*,*
  - ***boundary*** *(*$c_k \in B_k(X; G)$*) if* $c_k = \partial c_{k+1}$ *for some* $c_{k+1} \in C_{k+1}(X; G)$*.*
    Thus $Z_k(X)$ *is a subgroup of* $C_k(X)$*, and* $B_k(X)$ *is a subgroup of* $Z_k(X)$*.*
- *Two cycles* $c_k', c_k'' \in Z_k(X)$ *are called* ***equivalent*** *if* $c_k' - c_k''$ *is a boundary. This is an equivalence relation.*
- *The set* $H_k(X; G)$ *of equivalence classes is called the* k^{th} ***singular homology group****. Thus it is the factor group* $H_k(X) = Z_k(X)/B_k(X)$*.*

For Morse theory, one usually uses the field $\mathbb{R}$ for the additive group G. This way, $C_k(X)$ and analogously $Z_k(X)$, $B_k(X)$ and $H_k(X)$ will be $\mathbb{R}$-vector spaces. But whereas typically $C_k(X)$, $Z_k(X)$ and $B_k(X)$ are infinite dimensional, the homology group $H_k(X)$ is finite dimensional for compact manifolds X. By the universal coefficient theorem (Theorem 15.14 in [BT]) and the Theorem by de Rham (e.g.: Theorem 8.9 in [BT]), one even has $\dim(H_k(X)) = \mathrm{betti}_k(X)$, with the Betti numbers $\mathrm{betti}_k(X)$.

A continuous mapping of topological spaces $\varphi : X \to Y$ induces the homomorphism

$$\varphi_* : C_k(X; G) \to C_k(Y; G) \quad , \quad \varphi_*(c_k) = \sum_i g_i\,(\sigma_{k,i}, \varphi \circ f_i) \qquad (k \in \mathbb{N}_0)$$

of the chains, where $c_k := \sum_i g_i\ (\sigma_{k,i}, f_i)$. In turn, φ_* induces a homomorphism

$$\varphi_* : H_k(X; G) \to H_k(Y; G) \qquad (k \in \mathbb{N}_0). \tag{G.2.1}$$

This homomorphism only depends on the homotopy class of f. In particular, homotopy equivalent spaces have isomorphic singular homology groups.

G.10 Example The punctured plane and the circle are homotopy equivalent according to Example A.24.1. Therefore $H_k\big(\mathbb{C}\backslash\{0\}; G\big) = H_k\big(S^1; G\big)$. ◊

For the proof of the Morse inequalities, we will also need the *relative* homologies. If Y is a subspace of the topological space X, then the group $C_k(Y; G)$ is a subgroup of $C_k(X; G)$. The factor group will be denoted as

$$C_k(X, Y) := C_k(X)/C_k(Y)\,.$$

∂ maps $C_k(Y)$ into $C_{k-1}(Y)$ and thus defines a boundary operator

$$\partial : C_k(X, Y) \to C_{k-1}(X, Y)\,.$$

We define the group of the *relative cycles*

$$Z_k(X, Y) := \{c_k \in C_k(X, Y) \mid \partial c_k = 0\}$$

and the group of the *relative boundaries*

$$B_k(X, Y) := \big\{c_k \in C_k(X, Y) \mid \exists\, c_{k+1} \in C_{k+1}(X, Y) : \ c_k = \partial c_{k+1}\big\}\,.$$

The k^{th} *relative homology group* is the factor group

$$H_k(X, Y) := Z_k(X, Y)/B_k(X, Y)\,.$$

- As every cycle from $H_k(X)$ can be viewed as a cycle from $H_k(X, Y)$, we obtain a homomorphism

$$j : H_k(X) \to H_k(X, Y) \quad , \quad [z_k] \mapsto [z_k + C_k(Y)]\,.$$

- On the other hand, the inclusion $i : Y \to X$ with (G.2.1) induces a homomorphism

$$i_* : H_k(Y) \to H_k(X)\,.$$

- Finally we note that for $[z_k + C_k(Y)] \in H_k(X, Y)$ with $n \geq 1$, the cycle $z_k + C_k(Y)$ is a relative k-cycle, thus ∂z_k lies in $C_{k-1}(Y)$. Since $\partial\partial z_k = 0$ (Theorem G.8), ∂z_k is a $(k-1)$-cycle; thus $\partial z_k \in Z_{k-1}(Y)$ defines an element $[\partial z_k] \in H_{k-1}(Y)$. The corresponding mapping is

$$\partial_* : H_k(X, Y) \to H_{k-1}(Y) \quad , \quad [z_k + C_k(Y)] \mapsto [\partial z_k] \, .$$

G.11 Theorem *The sequence*

$$\ldots \xrightarrow{\partial_*} H_k(Y) \xrightarrow{i_*} H_k(X) \xrightarrow{j} H_k(X, Y) \xrightarrow{\partial_*} H_{k-1}(Y) \xrightarrow{i_*} \ldots H_0(X, Y) \to 0$$

is exact, i.e., $\ker(i_*) = \mathrm{im}(\partial_*)$ *,* $\ker(j) = \mathrm{im}(i_*)$ *, and* $\ker(\partial_*) = \mathrm{im}(j)$.

Proof: See Volume 3, §5 of Dubrovin, Fomenko and Novikov [DFN]. □

We return to the proof of the Morse inequalities. It is useful for this proof to compare the *Poincaré polynomial of the manifold* M^n,

$$P_M(t) := \sum_{\ell=0}^{n} \mathrm{betti}_\ell(M)\, t^\ell \, ,$$

with the *Poincaré polynomial of the Morse function* $f : M \to \mathbb{R}$,

$$Q_M(f, t) := \sum_{\ell=0}^{n} \mathrm{crit}_\ell(f)\, t^\ell \, .$$

By sufficiently small changes of f on disjoint neighborhoods of the critical points, we can achieve that the values $c_k := f(x_k)$ of the critical points $x_1, \ldots, x_N$ of f are pairwise different without changing their index. We number the c_k in increasing order, and define for $Y \subseteq X$ the relative Betti numbers

$$\mathrm{betti}_\ell(X, Y) := \dim H_\ell(X, Y) \qquad (\ell \in \mathbb{N}_0).$$

We choose regular values $a_i \in (c_i, c_{i+1})$, with $a_0 < c_1$ and $a_N > c_N$.

G.12 Lemma *The Poincaré polynomial of* f *is*

$$Q_M(f, t) = \sum_{k=1}^{N} \sum_{\ell=0}^{n} \mathrm{betti}_\ell\big(M_{a_k}, M_{a_{k-1}}\big)\, t^\ell \, .$$

Proof:

- From a property of the relative homology called *excision property*, it follows that $H_\ell(X, Y) = H_\ell(X/Y, *)$, where $*$ denotes a singleton and X/Y is the space with its quotient topology where Y is collapsed to a single point.
- In our case, for the critical point x_k of index m, the manifold $M_{a_k}/M_{a_{k-1}}$ is homotopy equivalent to $D^m/\partial D^m \cong S^m$, see the argument following the Morse Lemma G.6. Therefore, the Betti numbers satisfy

$$\mathrm{betti}_\ell\big(M_{a_k}, M_{a_{k-1}}\big) = \mathrm{betti}_\ell(S^m, *) = \delta(\ell, m)\,.\square$$

From the exactness of the sequence

$$\dots H_i \xrightarrow{f_i} H_{i+1} \xrightarrow{f_{i+1}} \dots \xrightarrow{f_{i+k-1}} H_{i+k} \xrightarrow{f_{i+k}} \dots$$

of finite dimensional vector spaces, i.e., $\dim\big(\ker(f_{\ell+1})\big) = \dim\big(\mathrm{im}(f_\ell)\big)$, we conclude with the dimension theorem $\dim(H_\ell) = \dim(\ker(f_\ell)) + \dim(\mathrm{im}(f_\ell))$ of linear algebra that the alternating sum satisfies:

$$\textstyle\sum_{\ell=i}^{i+k}(-1)^{\ell-i}\dim(H_\ell) \;=\; \dim\big(\ker(f_i)\big) + (-1)^k \dim\big(\mathrm{im}(f_{i+k})\big)\,.$$

Therefore, Theorem G.11 implies

$$\sum_{\ell=0}^{\infty}(-1)^\ell\,\big(\mathrm{betti}_\ell(Y) - \mathrm{betti}_\ell(X) + \mathrm{betti}_\ell(X, Y)\big) = 0\,. \tag{G.2.2}$$

Proof of the Morse Inequalities:

- Since $M = M_{a_N}$, Formula (G.1.2) for the Euler characteristic follows from (G.2.2) by summation over k, using $Y := M_{a_k}$ and $X := M_{a_{k-1}}$.
- In similar manner as (G.1.2), one proves the *strong* Morse inequalities

$$\textstyle\sum_{\ell=0}^{m}(-1)^{\ell-m}\,\mathrm{crit}_\ell(f) \geq \sum_{\ell=0}^{m}(-1)^{\ell-m}\,\mathrm{betti}_\ell(M) \qquad (m = 1, \dots, n-1). \tag{G.2.3}$$

The *weak* Morse inequalities (G.1.1) follow from this by adding (G.2.3) for pairs of adjacent values of m. □

G.3 Geodesic Motion and Morse Theory

In this section, we give an overview over applications and extensions of Morse theory, in particular for analyzing geodesics of a complete Riemannian manifold.

A geodesic on a Riemannian manifold (M, g) is a curve $c \in C^2(I, M)$ for which the parallel transport along c leaves the velocity vector c' invariant, in other words, which in local coordinates satisfies the geodesic equation (8.4).

For the interval $I = [0, 1]$, it is therefore an extremal of the energy functional (G.1.4) and of the *length functional*

$$\mathcal{L} : H^1(I, M) \to [0, \infty) \quad , \quad \mathcal{L}(c) := \int_I \sqrt{g_{c(t)}\big(\dot c(t),\, \dot c(t)\big)}\,\mathrm{d}t\,, \tag{G.3.1}$$

if compared with curves that have the same start and end points $c(0)$ and $c(1)$ respectively. An analogous statement can be made for arbitrary pairs of points on the geodesic.

If M is a submanifold of $\mathbb{R}^k$, and the Riemannian metric g on M is the restriction of the Euclidean metric of $\mathbb{R}^k$, then geodesics on M are distinguished by the fact that their acceleration $c''(t) \in T_{c(t)}\mathbb{R}^k$ is orthogonal to the tangent space $T_{c(t)}M$.

Geodesic motion, understood as a flow on the tangent bundle TM, is Hamiltonian with the Hamilton function

$$H : TM \to \mathbb{R} \quad , \quad v \mapsto \tfrac{1}{2} g_m(v, v) \ \text{ for } v \in T_m M \, . \tag{G.3.2}$$

Hereby, the symplectic form defining the Hamiltonian vector field X_H is the pullback from T^*M to TM of the canonical symplectic form ω_0 on T^*M under the bundle map $TM \to T^*M, \ v \mapsto g(v, \cdot)$. (See page 219.)

The length functional L allows to view a (connected) Riemannian manifold (M, g) as a metric space with the metric $d : M \times M \to [0, \infty)$,

$$d(q_0, q_1) := \inf \left\{ \mathcal{L}(c) \mid c \in H^1(I, M), \ c(0) = q_0, \ c(1) = q_1 \right\} . \tag{G.3.3}$$

The Cauchy-Schwarz inequality implies, for the time intervals $[t_0, t_1] \subseteq I$:

$$d\big(c(t_0), c(t_1)\big) \leq \sqrt{2\, \mathcal{E}(c)\ (t_1 - t_0)} \, .$$

G.13 Definition *The Riemannian manifold* (M, g)

- *is called* ***geodesically complete*** *if for every tangent vector* $v \in T_m M$*, there exists a geodesic* $c : \mathbb{R} \to M$ *with* $c(0) = m$ *and* $c'(0) = v$.
- *With these notations, on a geodesically complete manifold, the mapping*

$$\exp : TM \to M \quad , \quad v \mapsto c(1)$$

is called the ***exponential map****. Its restriction to* $T_m M$ *is called*

$$\exp_m : T_m M \to M \, .$$

This exponential map of differential geometry is not to be confused with the mapping (E.3.1) from the theory of Lie groups that bears the same name.[23]

G.14 Example

1. Compact manifolds are geodesically complete with respect to any Riemannian metric, because the flow-invariant level sets of the Hamilton function (G.3.2) will then always be compact. By Theorem 3.27, this implies the claim.

[23] If the Lie group G is equipped with a metric that is invariant both under the left and right action (E.1.3), the two notions coincide if the Lie algebra $\mathfrak{g}$ in $\exp : \mathfrak{g} \to G$ is viewed as the tangent space $T_e G$ of G at the neutral element $e \in G$. Examples are abelian Lie groups like $\mathbb{R}^n$ and the torus $\mathbb{T}^n$.

2. A simple example of an exponential map is the one for the round sphere $S^2 \subset \mathbb{R}^3$. Viewed from the tangent space $T_n S^2 \cong \mathbb{R}^2$ of the north pole n, the inverse of the tangent map $\exp_n : T_n S^2 \to S^2$ maps the (earth) sphere azimuthally to the plane. In this case, the geodesics are great circles.
   The north pole occurs at the origin and also again as concentric circles of radii $2\pi n$, $n \in \mathbb{N}$.
   In the figure on the right, one sees Antarctica with the south pole as a circle of radius π. $\Diamond$

Inverse exponential map of the earth at the north pole, namely an equidistant azimuthal projection.
[24]Image: courtesy of Wikipedia author RokerHRO

G.15 Theorem (Hopf and Rinow) *For a connected Riemannian manifold (M, g), the following statements are equivalent:*

1. *(M, g) is geodesically complete,*
2. *(M, d) (with the metric d from (G.3.3)) is a complete metric space,*
3. *the closed and (with respect to d bounded) subsets of M are compact.*

Under this hypothesis, there exists, for any two points q_0 and q_1 of M, a geodesic $c : [0, 1] \to M$ with $c(i) = q_i$ and with minimum length $\mathcal{L}(c) = d(q_0, q_1)$.

In order to investigate the geodesics by means of Morsetheory, one needs to understand their index with respect to the energy functional. This index is related to the conjugate points of the geodesics.

G.16 Definition

- *Let $c : [0, T] \to M$ be a geodesic segment on the Riemannian manifold (M, g). Then $q := c(t)$ is called a **conjugate point of** $p := c(0)$ (along c) if the linear mapping*

$$T_{c'(0)t} \exp_p \ : \ T_{c'(0)t} T_p M \ \longrightarrow \ T_q M$$

  has a nontrivial kernel.
- *The dimension of this kernel is called the **multiplicity** $\mathrm{Mult}_c(t)$ of the conjugate point.*

G.17 Remark (Conjugate Points)
So one is considering the linearization of the exponential map and looks whether a variation of the initial velocity of the geodesic starting at p will nevertheless lead

[24]This sum is finite!

to $q = \exp_p(c'(0)t)$ after a time t. The points $c(t)$ that are conjugate to p along the geodesic c will have time parameters that do not accumulate.

If q is a conjugate point to p along c, then p is also conjugate to q (along the same geodesic traversed in reverse).

In Example G.14.2, it is exactly the pairs of antipodes $(p, q) = (p, -p)$ and the pairs (p, p) of the sphere that are conjugate to each other, along each great circle segment meeting them. ◊

The Hilbert manifold $H^1(I, M)$ introduced in Remark G.3.2 has only the constant curves as critical points of the energy functional $\mathcal{E}$. However, by means of the end point mapping

$$\pi : H^1(I, M) \to M \times M \quad , \quad c \mapsto \big(c(0), c(1)\big) ,$$

one can define submanifolds of $H^1(I, M)$ and restrict $\mathcal{E}$ to these. Of particular importance are the spaces

$$\Omega_{p,q} M := \pi^{-1}\big((p, q)\big) \qquad (p, q \in M)$$

of curves that start at p and end at q, and

$$\Lambda M := \pi^{-1}\big(\Delta\big) \qquad \text{, with the diagonal } \Delta := \{(q, q) \mid q \in M\} .$$

G.18 Theorem (Index Theorem by Morse) *The index of the energy functional*

$$\mathcal{E} : \Omega_{p,q} M \longrightarrow [0, \infty)$$

at a geodesic c from p to q is $\sum_{t\in(0,1)} \mathrm{Mult}_c(t)$.

Proof: In Klingenberg [Kli2], *Theorem 2.5.9.* □

It is reasonable to view the space ΛM as the space of H^1*-loops* $c : S^1 \to M$. The energy functional restricted to ΛM (and again denoted as $\mathcal{E}$) is then invariant under the rotations $t \mapsto t + s$ on $S^1 = \mathbb{R}/\mathbb{Z}$. It is therefore always degenerate and thus is not a Morse function.[25] However, the critical points of $\mathcal{E} : \Lambda M \to [0, \infty)$ have finite index, and they are closed geodesics. The latter is shown by an integration by parts in formula (G.1.5) for the derivative $\mathrm{D}\mathcal{E}$, similar to the Hamiltonian variational principle in Theorem 8.16.

G.19 Theorem
If (M, g) is complete, then $\Omega_{p,q} M$ and ΛM are also complete metric spaces with respect to the metric on $H^1(I, M)$ induced by the scalar product (G.1.3).

---

[25]But it could be a *Morse-Bott function*: For these, the critical set is a closed submanifold and the Hessian in normal direction is nondegenerate.

Proof: See *Theorem 2.4.7* in KLINGENBERG [Kli2]. There it is assumed that M is compact. The slight generalization to complete (M, g) can be found, e.g., in *Proposition 4.1* of.[26] □

In order to find geodesics in $\Omega_{p,q}M$ or ΛM respectively as critical points of $\mathcal{E}$, one needs to check a certain condition called Palais-Smale condition. We generally will denote these spaces as Ω:

G.20 Definition *Ω fulfills the* **Palais-Smale Condition** *if all sequences $(c_k)_{k\in\mathbb{N}}$ of curves in Ω for which the sequence $\big(\mathcal{E}(c_k)\big)_{k\in\mathbb{N}}$ is bounded and which satisfy $\lim_{k\to\infty} \| \text{grad}\, \mathcal{E}(c_k)\| = 0$ have a convergent subsequence.*

G.21 Example (Violation of the Palais-Smale Condition)
If (M, g) is the (complete) Riemannian manifold $M := \mathbb{R}$ with the Euclidean metric g, then the Palais-Smale condition does not hold for ΛM.

For instance, the sequence of constant loops c_k with $c_k(t) := k$ satisfies the hypotheses in the definition, but it does not have a convergent subsequence. ♢

G.22 Lemma *The Palais-Smale condition holds for $\Omega_{p,q}M$ if (M, g) is complete; and it holds for ΛM if M is compact.*

Under the stated hypotheses, the gradient vector field grad $\mathcal{E}$ generates a complete flow on Ω. So one finds for instance in every connected component of $\Omega_{p,q}M$ a geodesic segment connecting p to q. These connected components can be indexed by the fundamental group $\pi_1(M)$.

But even in the case of a trivial fundamental group, one can frequently prove the existence of many such geodesic segments:

G.23 Example (Geodesics on Spheres)

- First let $M = S^n$ with the standard metric g and $p, q \in S^n$ not conjugate, i.e., not equal nor antipodes. We can connect these two points with infinitely many geodesic segments of multiplicities $k(n-1)$, $k \in \mathbb{N}_0$.
  However, these are all segments of *one* geodesic (lying in $\text{span}(p, q)$).
- If the metric is not the standard metric, one can again conclude, by means of Sard's theorem (see page 319) applied to the exponential map, that almost all $(p, q) \in M \times M$ are not conjugate, and again one obtains many connecting geodesic segments.
  But in general, these will now be geometrically different. ♢

Another question is about the existence and number of closed geodesics. One uses the energy functional $\mathcal{E} : \Lambda M \to [0, \infty)$ as introduced in Remark G.3.2 and restricted to the space of H^1-loops $c : S^1 \to M$.

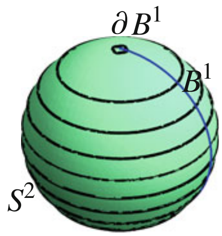
G.24 Theorem (Ljusternik-Fet) *On a* closed *(i.e., compact without boundary) Riemannian manifold, there exists a periodic geodesic.*

---

[26]M. Klein, A. Knauf: Classical Planar Scattering by Coulombic Potentials. LNP 13. Berlin: Springer, 1993.

Proof: Next to the original proof in [LF], see also *Theorem 3.7.7* of KLINGENBERG [Kli2], there is a different proof in [Kli2], *Theorem 2.4.20*.
Here is the idea of the proof:

- If the fundamental group $\pi_1(M)$ is nontrivial, then there are nontrivial conjugacy classes (see Theorem E.5) in $\pi_1(M)$, hence noncontractible loops. As generally, the connected components of ΛM correspond to the conjugacy classes, we find a closed geodesic by pulling such a loop taut by means of the gradient flow.
- However, if M is simply connected, then there exists some nontrivial homotopy group[27] $\pi_\ell(M)$, $2 \leq \ell \leq \dim(M)$. Assume $f : S^\ell \to M$ is not homotopic to a constant mapping. We decompose S^ℓ (in a way that is indicated in the figure for $\ell = 2$) into orbits of S^1, parametrized over $B^{\ell-1}$, where constant circles correspond to parameters from $\partial B^{\ell-1}$. We pull them taut together. Most loops shrink to constant loops. However, as f is not homotopic to a constant mapping, one loop converges to a closed geodesic of positive length. Example: In the case of $M = S^2$ with the standard metric and $f = \mathrm{Id}_{S^2}$, this geodesic is the equator. □



G.25 Remark (Existence of Many Closed Geodesics)
Typically, a closed Riemannian manifold (M, g) of $\dim(M) \geq 2$ has more than one closed geodesic. Here we call two geodesics geometrically different if their orbits in the unit tangent bundle $T_1 M$ are different. One can show this if the fundamental group $\pi_1(M)$ is sufficiently large (for instance for tori, see Theorem 8.33). But also the sphere S^2 with arbitrary Riemannian metric is an example where this has been proved, see BANGERT [Ban]. ◊

A major problem when applying Morse theory is the hypothesis that critical points be nondegenerate. While this hypothesis is true generically, it is difficult to verify in a single particular case.

The Lyusternik-Schnirelmann category is useful in this context, because it gives results that do not rely on nondegeneracy:

G.26 Definition
The ***Lyusternik-Schnirelmann category*** $\mathrm{cat}(X)$ *of a Hausdorff space* X *is the smallest cardinality of a family of contractible closed*[28] *sets* $A_i \subseteq X$ $(i \in I)$ *that cover* X*:*

$$X = \bigcup_{i\in I} A_i \,.$$

[27] This is a group of homotopy classes of maps $S^\ell \to M$ defined in analogy to the fundamental group.

[28] For our purposes, one can also use *open* covers, see R. Fox: On the Lusternik-Schnirelmann Category. The Annals of Mathematics, Second Series, **42**, 333–370 (1941).

G.27 Example (Circle)
The Lyusternik-Schnirelmann category $\mathrm{cat}(S^1)$ of the circle S^1 is 2. For instance, the two closed subsets $A_\pm := \{z \in S^1 \subset \mathbb{C} \mid \pm\Re(z) \le 1/2\}$ are contractible, and $A_- \cup A_+ = S^1$. But S^1 itself is not contractible. ♢

If X is compact, then $\mathrm{cat}(X) < \infty$. One has the result

G.28 Theorem *A function $f \in C^2(M, \mathbb{R})$ on a closed manifold M has at least* $\mathrm{cat}(M)$ *critical points.*

Proof:
A proof can be found in Volume 3, §19 of DUBROVIN, FOMENKO and NOVIKOV [DFN]. Roughly, the idea is as follows: ∇f, with respect to any Riemannian metric g, generates a complete gradient flow on M. We can write

$$M = \bigcup_{m \in \mathrm{Crit}(M)} W^s(m)\,,$$

where the stable manifolds $W^s(m)$ of different critical points are disjoint. From these $W^s(m)$, one can manufacture a covering in the sense of Definition G.26. □

We can often calculate the Lyusternik-Schnirelmann category by means of the notion of *cup length* (defined here for the coefficient ring $\mathbb{R}$):

G.29 Definition *The **cup length*** $\mathrm{cup}(M)$ *of a manifold M is the maximum number of elements $\alpha_1, \ldots, \alpha_p \in H^*(M)$ of degree ≥ 1 for which*

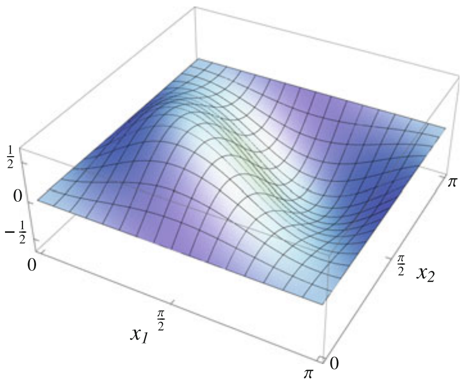
$$\alpha_1 \wedge \ldots \wedge \alpha_p \ne 0\,.$$

Consequently, the cup length of an n-dimensional manifold is at most n.

G.30 Theorem *For a compact manifold M,* $\mathrm{cat}(M) \ge \mathrm{cup}(M) + 1$.

Proof: In Volume 3, §19 of DUBROVIN, FOMENKO and NOVIKOV [DFN]. See also NICOLAESCU, [Ni, Theorem 2.58]. □

G.31 Example (Torus)
$\mathrm{cup}(\mathbb{T}^n) = n$, as can be seen by using a basis $\alpha_1, \ldots, \alpha_n \in H^1(\mathbb{T}^n)$ of $H^*(\mathbb{T}^n)$ (compare with Example B.54). Thus every smooth function $f : \mathbb{T}^n \to \mathbb{R}$ has at least $n + 1$ critical points.
This lower bound is sharp. In the example of the 2-torus $\mathbb{T}^2 := \mathbb{R}^2/\pi\mathbb{Z}^2$ with the function $f : \mathbb{T}^2 \to \mathbb{R}$,



$$x \mapsto \sin(x_1)\sin(x_2)\sin(x_1 + x_2)$$

(which is well-defined), f has exactly three critical points:

$(0, 0)$	(degenerate),
$(\pi/3, \pi/3)$	(maximum), and
$(2\pi/3, 2\pi/3)$	(minimum),

see the figure. ◊

"*Finishing a book is just like you took a child out in the yard and shot it.*"

TRUMAN CAPOTE

Appendix H
Solutions of the Exercises

H.1 Chapter 1, Introduction

Exercise 1.1 on page 6 (Kepler's Third Law):

(a) By the parametric equation (1.7), the minimal and maximal distances are

$$r_{\min} = R(\varphi_0) = \frac{p}{1+e} \quad \text{and} \quad r_{\max} = R(\varphi_0 + \pi) = \frac{p}{1-e}.$$

By definition, the major semiaxis $a = \frac{p}{1-e^2}$ is their arithmetic mean.
The distance between the two foci is $r_{\max} - r_{\min}$. Hence, using the gardener's definition of an ellipse,[29] the minor semiaxis $\tilde{b}$ satisfies, by Pythagoras, the relation

$$\tilde{b}^2 + (r_{\max} - r_{\min})^2/4 = (r_{\max} + r_{\min})^2/4\,, \quad \text{or} \quad \tilde{b} = (r_{\min}\, r_{\max})^{1/2} = b.$$

(b) The area of the ellipse is πab, hence by Kepler's second law equal to $\ell T/2$, where T is the orbital period. Hence $T = 2\pi ab/\ell$.

(c) With $b = p/\sqrt{1-e^2} = \sqrt{pa} = \ell\sqrt{a/\gamma}$, one obtains from part b for the orbital period:

$$T = \tfrac{2\pi ab}{\ell} = 2\pi \tfrac{a^{3/2}}{\sqrt{\gamma}}. \qquad \square$$

[29]Namely, the description of an ellipse as the set of those points for which the sum of the distances to the two foci is constant; so an ellipse can be drawn by attaching a twine to two nails and pulling it taut with a pencil. Elliptic flower beds were popular during the Baroque and Rococo periods.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

H.2 Chapter 2, Dynamical Systems

Exercise 2.5 on page 13 (Cantor Set):

- The Cantor set is defined as $\tilde{C} := \bigcap_{n\in\mathbb{N}} C_n$ with $C_0 := I$, where $C_{n+1} \subset C_n$ is obtained from C_n by removing the middle third of each interval.
- Since $f(x) = 3x$ for $x \in [0, 1/3]$, and $f(x) = 3(1 - x)$ for $x \in [2/3, 1]$, it follows that $f(C_{n+1}) = C_n$, hence $f(\tilde{C}) = \tilde{C}$ and thus $\tilde{C} \subseteq C$.
- Conversely, if $x_0 \in I \setminus \tilde{C}$, then there exists exactly one $n \in \mathbb{N}_0$ with $x_n \in C_n$, but $x_{n+1} \notin C_{n+1}$. Then $x_{n+1} \in L$. □

Exercise 2.12 on page 15 (Period):

1. • For $\Phi_t(m) = \exp(2\pi\imath t\alpha)m$, one has $\Phi_0(m) = \exp(0)m = m$ and $\Phi_{t_1} \circ \Phi_{t_2}(m) = \exp(2\pi\imath t_1\alpha)\exp(2\pi\imath t_2\alpha)m = \exp(2\pi\imath(t_1 + t_2)\alpha)m = \Phi_{t_1+t_2}(m)$.
   So $(\Phi_t)_{t\in\mathbb{Z}}$ is a dynamical system on S^1.
   • For $\alpha = q/p$ with $q \in \mathbb{Z},\ p \in \mathbb{N}$, one has $\Phi_p(m) = \exp\left(2\pi\imath p\frac{q}{p}\right) m = m$, so p is a period of $m \in S^1$.
   The minimal period $r \in \mathbb{N}$ of m must divide p (in formulas: $r|p$), and $\Phi_r(m) = m$ implies $\exp\left(2\pi\imath\frac{rq}{p}\right) = 1$, hence $p|rq$. As q and p were assumed to be relatively prime, it follows that $p|r$. Since $r|p$ also holds, we conclude $r = p$.
   • If $\Phi_t(m) = m$ for some $t \in \mathbb{Z}$ and $m \in S^1$, then $\exp(2\pi\imath\alpha t) = 1$, and thus $\alpha t \in \mathbb{Z}$. If $\alpha \in \mathbb{R}\backslash\mathbb{Q}$, then $t = 0$.
2. By induction, we obtain from $m, n \in \mathbb{N}$ and $f_m(z) := z^m$ that $f_m^{(n)} = f_{m^n}$.
   For $\ell \in \mathbb{N}$, the equation $f_\ell(z) = z$ implies $z^\ell = z$, or $z^{\ell-1} - 1 = 0$. For $\ell > 1$, this polynomial has degree $\ell - 1$, and thus it has exactly $\ell - 1$ zeros on the unit circle in $\mathbb{C}$.
   Therefore $P_n(f_m) = P_1\left(f_m^{(n)}\right) = P_1\left(f_{m^n}\right) = m^n - 1$. □

Exercise 2.15 on page 16 (Minimal Period of a Dynamical System):
For dynamical systems, the mappings $\Phi_t : M \to M$ are bijective. If M is a finite set and $m \in M$, then there exist times $t_1 < t_2$ with $\Phi_{t_1}(m) = \Phi_{t_2}(m)$. But then $t := t_2 - t_1$ is a period: $\Phi_t(m) = m$. Hence every point $m \in M$ is periodic, and as $t \in \mathbb{Z}$, there exists a minimal period $T(m) \in \mathbb{N}$.

As the least common multiple $\hat{T} \in \mathbb{N}$ of the $T(m)$ $(m \in M)$ exists (since M is finite), all $m \in M$ satisfy $\Phi_{\hat{T}}(m) = m$. On the other hand, $\hat{T}$ is also the minimal period of the dynamical system. □

Exercise 2.19 on page 19 (Shift):

1. $d_\mathcal{A}$ is a metric on $\mathcal{A}$, therefore so is $2^{-|j|}d_\mathcal{A}$, for arbitrary $j \in \mathbb{Z}$.
   Also, since $\sum_{j\in\mathbb{Z}} 2^{-|j|} = 3 < \infty$ is finite, d is a product metric on $\mathcal{A}^\mathbb{Z}$:

   - For $x \neq y$, there exists $j \in \mathbb{Z}$ with $x_j \neq y_j$, hence $d(x, y) \geq 2^{-|j|} > 0$.
   - $d(x, y) = d(y, x)$, since $d_\mathcal{A}(a, b) = d_\mathcal{A}(b, a)$.

- $d(x,z) = \sum_{j\in\mathbb{Z}} 2^{-|j|} d_{\mathcal{A}}(x_j, z_j) \le \sum_{j\in\mathbb{Z}} 2^{-|j|}\big(d_{\mathcal{A}}(x_j, y_j) + d_{\mathcal{A}}(y_j, z_j)\big) = d(x,y) + d(y,z)$.
  Φ is continuous, because $d\big(\Phi_{\pm 1}(x), \Phi_{\pm 1}(y)\big) \le 2d(x,y)$.

2. - $m = (m_j)_{j\in\mathbb{Z}}$ is n-periodic if and only if $m_{j+kn} = m_j$ for $j = 0, \ldots, n-1$ and $k \in \mathbb{Z}$. So there are exactly 2^n points with period n.
   - The minimal period T of an n-periodic point $m \in M$ divides n (see Theorem 2.13). Therefore, for $n = 2,\ 3,\ 4$, there are exactly $2 = 2^2 - 2,\ 6 = 2^3 - 2,\ 12 = 2^4 - 2 - 2$ points, respectively, with minimal period n.
   - The periodic orbits with minimal period k comprise k points. Therefore, there are $2 = 2/1,\ 1 = 2/2,\ 2 = 6/3,\ 3 = 12/4$ orbits, respectively, with minimal periods $1, 2, 3, 4$. So there are exactly $2, 3, 4, 6$ orbits with these respective periods.
     If $B(n)$ denotes the number of periodic orbits with minimal period $n \in \mathbb{N}$, we have
     $$2^n = \sum_{d:\, d|n} d\ B(d)\,,$$
     and the Möbius inversion of this relation tells us that
     $$B(n) = \frac{1}{n} \sum_{d:\, d|n} 2^d \mu(\tfrac{n}{d})\,.$$
     Here, $\mu\colon \mathbb{N} \to \{-1, 0, 1\}$ is the *Möbius function*
     $$\mu(n) := \begin{cases} 0 & \text{if } n \text{ has a repeated prime divisor,} \\ 1 & \text{if } n = 1 \text{ or } n \text{ has an even number of (distinct) prime divisors,} \\ -1 & \text{if } n \text{ has an odd number of (distinct) prime divisors} \end{cases}$$
     If n is prime, the formula for B simplifies to $B(n) = \frac{1}{n}\big(2^n - 2\big)$.
3. Now take for instance $x = (x_j)_{j\in\mathbb{Z}}$ with $x_j := 0$ for $j \le 0$ and $(x_j)_{j\in\mathbb{N}}$ the sequence
   0 1 00 01 10 11 000 001 010 011 . . ., which is obtained by concatenating the bit sequences of lengths 1, 2, . . . lexicographically. □

Exercise 2.22 on page 20 (Stability):

1. The maps $\Phi_t : \mathbb{C} \to \mathbb{C},\ \Phi_t(m) := \lambda^t m\ (t \in \mathbb{Z})$ for $\lambda \in \mathbb{C}\backslash\{0\}$ form a continuous dynamical system, because the Φ_t are linear, hence continuous, $\Phi_0(m) = \lambda^0 m = m$, and $\Phi_{t_1} \circ \Phi_{t_2}(m) = \lambda^{t_1}\lambda^{t_2} m = \lambda^{(t_1+t_2)} m$ holds true. Due to linearity, 0 is a fixed point.
2. The neighborhood basis $\{U_\varepsilon(0) \mid \varepsilon > 0\}$ of 0 satisfies $\Phi_t\big(U_\varepsilon(0)\big) \subseteq U_\varepsilon(0)$ for all $t \ge 0$ exactly if $|\lambda| \le 1$. So 0 is Lyapunov-stable for those values of λ.
   In contrast, if $|\lambda| > 1$, then $\lim_{t\to+\infty} |\lambda|^t \varepsilon = \infty$, and because $\Phi_t\big(U_\varepsilon(0)\big) = U_{|\lambda|^t\varepsilon}(0)$, the point 0 is not Lyapunov-stable in this case.

3. If $|\lambda| < 1$, then 0 is not only Lyapunov-stable, but also the radius $|\lambda|^t\varepsilon$ of the disc $\Phi_t\big(U_\varepsilon(0)\big)$ tends to 0 as $t \to +\infty$. Hence 0 is asymptotically stable. □

Exercise 2.25 on page 21 (Attractor):

(a) The union A of two attractors A_1, A_2 is again an attractor, because, given forward invariant neighborhoods $U_i \subseteq M$ of A_i, their union $U := U_1 \cup U_2$ is a forward invariant neighborhood of A, and given an open neighborhood $V \subseteq U$ of A, the $V_i := V \cap U_i \subseteq U_i$ are such neighborhoods for A_i. Then there exist $\tau_i > 0$ with $\Phi_t(U_i) \subseteq V_i$ for all $t \geq \tau_i$. Letting $\tau := \max(\tau_1, \tau_2)$, it then follows that $\Phi_t(U) \subseteq V$ for all $t \geq \tau$.

(b) $A \subseteq \bigcap_{t\geq 0} \Phi_t(U_0)$ follows from $A \subseteq U_0$ and $\Phi(t, A) = A$ for all $t \in G$. Let $x \in \bigcap_{t\geq 0} \Phi_t(U_0) \setminus A$. Then $V := U_0 \setminus \{x\}$ is open and satisfies $A \subseteq V \subseteq U_0$. So there exists $\tau \geq 0$ with $\Phi(t, U_0) \subseteq V$ for all $t \geq \tau$. This contradicts $x \in \bigcap_{t\geq 0} \Phi(t, U_0)$. □

Exercise 2.27 on page 23 (Logistic Family):

1. • For $f_4(x) = 4x(1-x)$, one has $f_4\left(\frac{1}{2}\right) = 1$, the maximum, and $f(0) = f(1) = 0$. As $f_4^{(n)} - \mathrm{Id}$ is a polynomial of degree 2^n (degrees multiply when polynomials are composed), it can have at most 2^n zeros. Hence there are at most 2^n fixed points of $f_n^{(n)}$.
   • On the other hand, for $f_4^{(n)}$ there are points $x_k^{(n)}$ $(k = 0, \ldots, 2^n)$ with

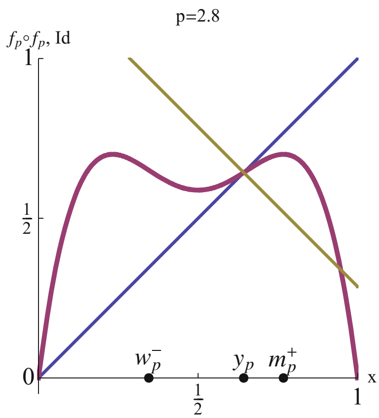
$$x_0^{(n)} = 0 \quad , \quad x_k^{(n)} < x_{k+1}^{(n)} \quad \text{and} \quad x_{2^n}^{(n)} = 1\,,$$

   such that $f_4^{(n)}\left(x_{2\ell}^{(n)}\right) = 0$ and $f_4^{n}\left(x_{2\ell+1}^{(n)}\right) = 1$.
   This follows by induction from $x_1^{(1)} := \frac{1}{2}$, because f_4 is strictly increasing on $\left[0, \frac{1}{2}\right]$, from 0 to 1, and then strictly decreasing on $\left[\frac{1}{2}, 1\right]$ back to 0. Therefore there are at least 2^n fixed points of $f_4^{(n)}$.

2. Now $f_p(y_p) = y_p$, and for $p \geq 1$, one has $y_p \in [0, 1]$. Thus in the range $p \in (1, 3)$ of parameters, y_p is the second fixed point of f_p, next to 0. For parameters $p \in (1, 2]$, it was shown in Example 2.26 that $\lim_{n\to\infty} f_p^{(n)}(x) = y_p$ for all $x \in (0, 1)$. For $p \in (2, 3)$, we consider

$$f_p^{(2)}(x) = p^2 x(1-x)\big(1 - px(1-x)\big)\,.$$

   Let's start by analyzing this function.



- For these parameters, $\frac{1}{2}$ is a minimum of $f_p^{(2)}$, and the maxima of $f_p^{(2)}$ are at $m_p^\pm := \frac{1}{2} \pm \frac{\sqrt{p(p-2)}}{2p}$. The maximum value is $f_p^{(2)}(m_p^\pm) = \frac{p}{4}$.
- The two points of inflection of $f_p^{(2)}$ are at $w_p^\pm := \frac{1}{2} \pm \frac{\sqrt{p(p-2)/3}}{2p}$, and $\frac{d}{dx} f_p^{(2)}(w_p^\pm) = \pm\left(\frac{p(p-2)}{3}\right)^{3/2}$.

  As the interval $(2, 3)$ lies inside the interval $\left(2, 1+\sqrt{5}\right)$, for which the inequality $f_p^{(2)}(m_p^+) < m_p^+$ holds, the interval $[m_p^+, 1)$ is mapped into $(0, m_p^+)$ by $f_p^{(2)}$, and also $f_p^{(2)}\big((0, m_p^+)\big) \subseteq (0, m_p^+)$.

  We show that the distance of x to the fixed point y_p decreases under iteration. This follows from the estimate

$$|f_p^{(2)}(x) - y_p| < |x - y_p|\,,$$

  which is valid for $x \in (0, m_p^+),\ x \neq y_p$. After all, the graph of $f_p^{(2)}$ lies above the diagonal in $(0, y_p)$, but also below the line $x \mapsto 2y_p - x$ through the fixed point, because $\frac{d}{dx} f_p^{(2)}(x) \geq \frac{d}{dx} f_p^{(2)}(w_p^-) > -1$. An analogous argument applies to the interval (y_p, m_p^+).

  An approach that applies to all $p \in (1, 3]$ can be found in DENKER [De], Chapter 1.5. □

Exercise 2.30 on page 23 (Conjugacy): Let $h \circ \Phi_t^{(1)} = \Phi_t^{(2)} \circ h$ $(t \in G)$, for a homeomorphism $h : M^{(1)} \to M^{(2)}$. Since conjugacy is an equivalence relation, it suffices to show one implication for each of the following equivalencies.

(a) $x \in M^{(i)}$ is an equilibrium if and only if $\Phi_t^{(i)}(x) = x$ $(t \in G)$. If $x_1 \in M^{(1)}$ is an equilibrium, then it follows

$$\Phi_t^{(2)}(x_2) = \Phi_t^{(2)} \circ h(x_1) = h \circ \Phi_t^{(1)}(x_1) = h(x_1) = x_2\,.$$

If $U^{(2)} \subseteq M^{(2)}$ is a neighborhood of x_2, then $U^{(1)} := h^{-1}(U^{(2)}) \subseteq M^{(1)}$ is one for x_1. If x_1 is Lyapunov-stable, then there exists a neighborhood $V^{(1)} \subseteq U^{(1)}$ of x_1 with $\Phi_t^{(1)}(V^{(1)}) \subseteq U^{(1)}$ $(t \geq 0)$. Accordingly, $V^{(2)} := h(V^{(1)}) \subseteq U^{(2)}$ is a neighborhood of x_2 with

$$\Phi_t^{(2)}\big(V^{(2)}\big) = \Phi_t^{(2)} \circ h\big(V^{(1)}\big) = h \circ \Phi_t^{(1)}\big(V^{(1)}\big) \subseteq h\big(U^{(1)}\big) = U^{(2)}\,.$$

The asymptotic stability carries over analogously.

Parts (b) and (c) are routine. □

Exercise 2.45 on page 29 (Diffeomorphism Group):

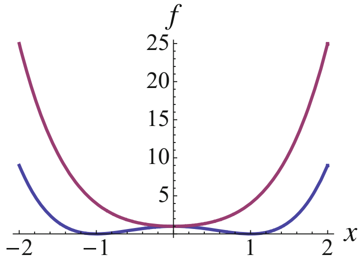
- Let (U, φ) be a coordinate chart of M with $x \in U$, and let $V := \varphi(U) \subseteq \mathbb{R}^n$ denote the image of that chart. For sufficiently small $\varepsilon > 0$, the ε-ball $U_\varepsilon(\tilde{x})$ about $\tilde{x} := \varphi(x)$ is entirely contained in V. Let $\tilde{\chi} \in C^\infty\big(U_\varepsilon(\tilde{x}), [0, 1]\big)$ be a cutoff function, say, for instance, $\tilde{\chi}\restriction_{U_{\varepsilon/4}}(\tilde{x}) = 1$ and $\tilde{\chi}(z) = 0$ for $z \notin U_{\varepsilon/2}(\tilde{x})$. Now if $\tilde{y} \in U_{\varepsilon/4}(\tilde{x})$, then $\tilde{v} : U_\varepsilon(\tilde{x}) \to \mathbb{R}^n$, $\tilde{v}(z) := \tilde{\chi}(z) \cdot (\tilde{y} - \tilde{x})$ defines a vector field that vanishes outside $U_{\varepsilon/2}(\tilde{x})$ and equals $\tilde{y} - \tilde{x}$ inside $U_{\varepsilon/4}(\tilde{x})$. Extending the lift of the vector field to $\varphi^{-1}\big(U_\varepsilon(x)\big) \subseteq U$ by 0, we get a vector field v on M. Its time-1 flow $f \in \mathrm{Diff}(M)$ exists (by an argument that is analogous to the proof of Theorem 3.27). Moreover, with $y := \varphi^{-1}(\tilde{y})$, one has $f(x) = y$, because within the chart, all $z_t := (1-t)\tilde{x} + t\tilde{y}$ satisfy $\tilde{v}(z_t) = \tilde{y} - \tilde{x}$.

- Therefore, the set $M_x := \{y \in M \mid \text{ there is } f \in \mathrm{Diff}(M) \text{ with } f(x) = y\}$ is open and non-empty.
  But $M \backslash M_x$ is also open, for if $z \in M\backslash M_x$ cannot be reached, then all of M_z cannot be reached either; this uses that $\mathrm{Diff}(M)$ is a group. But by assumption, M is connected; therefore $M_x = M$. □

H.3 Chapter 3, Ordinary Differential Equations

Exercise 3.12 on page 37 (Single Differential Equations of First Order):

1. - $f_1(x) = 0$ if and only if $|x| = 1$, otherwise $f_1(x) > 0$. Therefore the minimal invariant sets are $(-\infty, -1)$, $\{-1\}$, $(-1, 1)$, $\{1\}$, and $(1, +\infty)$. The solutions in the open intervals are strictly increasing.
     As $f_1(x) \geq x^4/2$ provided $|x| \geq 2$, the intervals of existence are bounded above for initial values $x_0 > 1$, and bounded below for $x_0 < -1$.



   - One has $f_2(x) \geq f_2(0) = 1$. This makes the entire phase space $\mathbb{R}$ the only non-empty invariant set; in particular, there are no fixed points.
     The solutions are strictly increasing and exist only for a finite interval of time, because $f_2(x) \geq x^4$.

2. The differential equation $\dot{x} = f(x)$ for $f : (0, \infty) \to \mathbb{R}$, $x \mapsto x^\alpha$ has the solutions $x(t) = e^t x_0$ if $\alpha = 1$, and $x(t) = \left(\beta t + x_0^\beta\right)^{1/\beta}$ with $\beta := 1 - \alpha$ otherwise. The latter solution is calculated for the initial value $x_0 > 0$ by separation of variables: $\big(x(t)^\beta - x_0^\beta\big)/\beta = \int_{x_0}^{x(t)} y^{-\alpha} \mathrm{d}y = \int_0^t \mathrm{d}s = t$.
   So, whereas in the linear case $\alpha = 1$, the solution exists for all times, it exists only for the time interval $\left(-\infty\,,\ x_0^\beta/|\beta|\right)$ in case $\alpha > 1$, and only for $t \in \left(-x_0^\beta/\beta\,,\ +\infty\right)$ in case $\alpha \in [0, 1)$. □

Exercise 3.19 on page 42 (Picard-Lindelöf):

(a) The maximal time guaranteed by Theorem 3.17 is $\varepsilon(r) := \min\{r, \frac{r}{N(r)}, \frac{1}{2L(r)}\}$ with $r > 0$, $B_r(0) \subseteq D_f = \mathbb{R}$, $N(r) := \max\{|f(x)| \mid x \in B_r(0)\} = e^r$ and the Lipschitz constant $L(r) := \mathrm{Lip}\big(f \restriction B_r(0)\big) = e^r$.
For $r = \frac{1}{2}$, one obtains $\varepsilon(r) = \frac{1}{2\sqrt{e}} \approx 0.303$, which is the maximal value.

(b) The Picard iteration is set up with this recursion formula:

$$x_0(t) := x_0 = 0 \quad \text{and} \quad x_{j+1} := x_0 + \int_{t_0}^{t} f\big(x_j(\tau)\big)\, \mathrm{d}\tau = \int_0^t e^{-x_j(\tau)}\, \mathrm{d}\tau.$$

(c) The maximal solution of the IVP is $\varphi\colon (-1, \infty) \to \mathbb{R}$, $\varphi(t) = \log(1+t)$. □

Exercise 3.25 on page 46

The sine function is Lipschitz continuous on $\mathbb{R}$ (but not on $\mathbb{C}$!). Therefore, the initial value problem $\dot{x} = \sin x,\ x_0 = \pi/2$ has a unique solution. It is obtained by separating variables:

$$t = \int_{x_0}^{x} \frac{1}{\sin y}\, \mathrm{d}y = \log\left(\tan\left(\frac{y}{2}\right)\right)\Big|_{x_0}^{x} = \log\left(\frac{\tan(x/2)}{\tan(\pi/4)}\right),$$

hence $x(t) = 2\arctan(e^t)$. Therefore, $\lim_{t\to+\infty} x(t) = \lim_{y\to+\infty} 2\arctan(y) = \pi$ and $\lim_{t\to-\infty} x(t) = 2\arctan(0) = 0$. □

Exercise 3.28 on page 47 (Existence of the Flow):

(a) For an arbitrary number $n \in \mathbb{N}$ of degrees of freedom, $H : \mathbb{R}^{2n} \to \mathbb{R}$, $H(x) := \|x\|^2$ is a function whose sublevel sets $H^{-1}\big((-\infty, E]\big)$ are compact balls. H is the Hamilton function of a harmonic oscillator.

(b)
- H is a constant of motion because a solution $\varphi\colon [-\varepsilon, \varepsilon] \to \mathbb{R}^{2n}$ to the Hamiltonian differential equation satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} H\big(\varphi(t)\big) = \sum_{j=1}^{2n} \frac{\partial H}{\partial x_j}\big(\varphi(t)\big)\, \dot{\varphi}_j(t) = 0\,.$$

- We consider the sublevel set $P_E := H^{-1}\big((-\infty, E]\big)$ for $E > 0$. It is invariant under the Hamiltonian flow, because H is a constant of motion. Since P_E is compact by hypothesis, the vector field $\mathbb{J}\, \nabla H \restriction_{P_E}$ with $\mathbb{J} = \left(\begin{smallmatrix} 0 & -\mathbb{1}_n \\ \mathbb{1}_n & 0 \end{smallmatrix}\right)$ is Lipschitz continuous by Lemma 3.14. The Hamiltonian flow on P_E is therefore defined for all times. As E was arbitrary and $\mathbb{R}^{2n} = \bigcup_{E>0} P_E$, the completeness of the vector field on $\mathbb{R}^{2n}$ follows. □

Exercise 3.30 on page 48:

The linear differential equation $\ddot{x} = x$ has the two solutions $x_\pm : \mathbb{R} \to \mathbb{R},\ x_\pm(t) = \exp(\pm t)$ as a basis of the solution space, specifically $x_\pm(0) = 1,\ \dot{x}_\pm(0) = \pm 1$. For the initial conditions $(x_0, \dot{x}_0) = (1, -1)$, one obtains therefore the solution $x(t) = x_-(t) = \exp(-t)$, with $0 = \lim_{t\to\infty} x(t)$. □

Exercise 3.41 on page 53 (Escape time):
The Hamiltonian differential equations are, as claimed,

$$\dot{p} = -\frac{\partial H}{\partial q} = -\frac{m}{q^2} \quad , \quad \dot{q} = \frac{\partial H}{\partial p} = p\,.$$

We have the conserved quantity $H(p, q) = \frac{1}{2}p^2 - \frac{m}{q} = E$, hence

$$q = \frac{m}{\frac{1}{2}p^2 - E}\,, \quad \text{and thus} \quad p = \pm\sqrt{2\big(E + \tfrac{m}{q}\big)}\,.$$

We give a combined solution for parts (a) and (b):

- $p_0 > 0$, $E := H(p_0, q_0) \geq 0$: In this case,

  $$\dot{q}(t) = \sqrt{2\big(E + \tfrac{m}{q(t)}\big)} \geq \sqrt{2E} \;\text{, hence for } t \geq 0:\; q(t) \geq q_0 + t\sqrt{2E}\,.$$

  So the left boundary of the domain $(0, \infty)$ of q is never reached for positive times. Moreover, we have for $t \geq 0$ and again $p_0 > 0$, $E \geq 0$:

  $$\dot{q}(t) = \sqrt{2\big(E + \tfrac{m}{q(t)}\big)} \leq \sqrt{2(E + \tfrac{m}{q_0})} \implies q(t) \leq q_0 + t\sqrt{2(E + \tfrac{m}{q_0})}\,.$$

  This means that for $p_0 > 0$ and $E \geq 0$, the escape time is $T^+(p_0, q_0) = +\infty$.
- $p_0 \leq 0$, $E := H(p_0, q_0) < 0$. By separation of variables, the differential equation $\dot{q} = -\sqrt{2\big(\frac{m}{q} - |E|\big)}$ becomes

  $$t = -\int_{q_0}^{q} \frac{\mathrm{d}\tilde{q}}{\sqrt{2\big(\frac{m}{\tilde{q}} - |E|\big)}} = g(q) - g(q_0)$$

  with $g(q) = (2|E|)^{-\frac{3}{2}}\big(q\sqrt{2|E|}\sqrt{2(\frac{m}{q} - |E|)} + m\arcsin(1 - 2|E|\frac{q}{m})\big)$; this is defined for $q \in (0, \frac{m}{|E|}]$. So for $q \searrow 0$, we obtain the escape time $T^+(p_0, q_0) = \lim_{q\searrow 0}(g(q) - g(q_0)) = \frac{m\pi}{2(2|E|)^{\frac{3}{2}}} - g(q_0)$.
- $p_0 > 0$, $E < 0$. The differential equation for q reads $\dot{q} = \sqrt{2\big(\frac{m}{q} - |E|\big)}$ and can be rewritten as

  $$t = \int_{q_0}^{q} \frac{\mathrm{d}\tilde{q}}{\sqrt{2\big(\frac{m}{\tilde{q}} - |E|\big)}} = g(q_0) - g(q)\,.$$

The differential equation loses its validity at $0 = \dot{q} = \sqrt{2\left(\frac{m}{q} - |E|\right)}$, namely when $q = \frac{m}{|E|}$. Until this point, the time $g(q_0) - g(\frac{m}{|E|}) = g(q_0) + \frac{m\pi}{2(2|E|)^{\frac{3}{2}}}$ elapses; thereafter the parameter range $p_0 \le 0$, $E < 0$ takes over. So we obtain the escape time $g(q_0) + \frac{3m\pi}{2(2|E|)^{\frac{3}{2}}}$.

- $p_0 \le 0$, $E = 0$. By separation of variables, the differential equation $\dot{q} = -\sqrt{2\frac{m}{q}}$ becomes

$$t = -(2m)^{-\frac{1}{2}} \int_{q_0}^{q} \sqrt{\tilde{q}}\, \mathrm{d}\tilde{q} = -\frac{2}{3\sqrt{2m}}\left(q^{\frac{3}{2}} - q_0^{\frac{3}{2}}\right) \iff q = \left(q_0^{\frac{3}{2}} - 3t\sqrt{\tfrac{m}{2}}\right)^{\frac{2}{3}}.$$

We can read off the escape time $\frac{\sqrt{2}}{3\sqrt{m}} q_0^{\frac{3}{2}}$ directly.

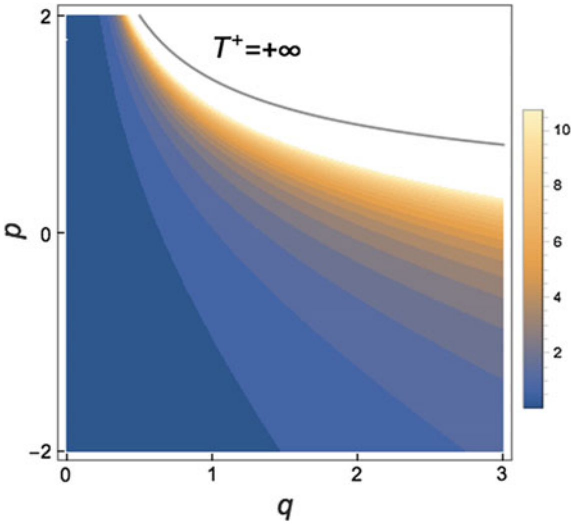


Figure H.1 Contour plot of escape time for the (unregularized) Kepler problem

- $p_0 < 0$, $E > 0$. The differential equation now reads $\dot{q} = -\sqrt{2\left(\frac{m}{q} + E\right)}$, and this leads to

$$t = -\int_{q_0}^{q} \frac{\mathrm{d}\tilde{q}}{\sqrt{2\left(\frac{m}{\tilde{q}} + E\right)}} = g(q) - g(q_0) \quad \text{with}$$

$$g(q) = \frac{\sqrt{2}}{4E}\left(\frac{m}{\sqrt{E}} \log\left(m + 2Eq + 2q\sqrt{E}\sqrt{\tfrac{m}{q} + E}\right) - 2q\sqrt{\tfrac{m}{q} + E}\right) \quad (q > 0).$$

With $\lim_{q \searrow 0} g(q) = m\log(m)/(2E)^{\frac{3}{2}}$, the escape time is

$$T^+(p_0, q_0) = g(q_0) + m\log(m)/(2E)^{\frac{3}{2}}\,.$$

Figure H.1 shows a contour plot of T^+.

(c) One example is $\dot{x} = x^2$. The solution to the IVP with $x(0) = x_0$ is unique, and $x(t) = 0$ is a solution. For $x_0 \in \mathbb{R} \setminus \{0\}$, the solution is (see Exercise 3.12) $x(t) = \left(x_0^{-1} - t\right)^{-1}$ on $\left(x_0^{-1}, \infty\right)$ for $x_0 < 0$; or on $\left(-\infty, x_0^{-1}\right)$ for $x_0 > 0$. □

H.4 Chapter 4, Linear Dynamics

Exercise 4.11 on page 67 (Matrix Exponential):
$A = \mathbb{1} + B$ with $B := \begin{pmatrix} 0&0&0\\ 1&0&0\\ 1&1&0 \end{pmatrix}$. So the solution operator for all times $t \in \mathbb{R}$ is

$$\exp(At) = e^t \exp(Bt) = e^t\left(\mathbb{1} + Bt + \tfrac{1}{2}B^2t^2\right) = e^t \begin{pmatrix} 1&0&0\\ t&1&0\\ t+t^2/2&t&1 \end{pmatrix}.$$ □

H.5 Chapter 5, Classification of Linear Flows

Exercise 5.5 on page 83 (Index): 1 is a double eigenvalue, and -1 a single eigenvalue of the matrix $A := \begin{pmatrix} 1&1&-1\\ 0&-1&2\\ -2&-1&1 \end{pmatrix}$. Thus the matrix is hyperbolic and its index is 1. Now $B^{-1}AB = J$ with

$$B = \begin{pmatrix} -1&-1&1\\ 2&0&-2\\ 2&1&0 \end{pmatrix},\; B^{-1} = \tfrac{1}{4}\begin{pmatrix} 2&1&2\\ -4&-2&0\\ 2&-1&2 \end{pmatrix} \text{ and Jordan matrix } J = \begin{pmatrix} 1&1&0\\ 0&1&0\\ 0&0&-1 \end{pmatrix}.$$

Thus a fundamental system of solutions is

$$\mathbb{R} \ni t \mapsto e^t\begin{pmatrix} -1\\ 2\\ 2 \end{pmatrix} \;,\; \mathbb{R} \ni t \mapsto e^t\left(t\begin{pmatrix} -1\\ 2\\ 2 \end{pmatrix} + \begin{pmatrix} 1\\ 0\\ -1 \end{pmatrix}\right) \;,\; \mathbb{R} \ni t \mapsto e^{-t}\begin{pmatrix} 1\\ -2\\ 0 \end{pmatrix},$$

and the latter among them remains bounded as $t \to +\infty$. □

Exercise 5.12 on page 89: (Hooke's Law)
We have $(At)^2 = at^2\begin{pmatrix} 1&0\\ 0&1 \end{pmatrix}$, hence

$$\exp(At) = \sum_{n=0}^{\infty}\frac{(At)^n}{n!} = \mathbb{1}\sum_{k=0}^{\infty}\frac{(at^2)^k}{(2k)!} + At\sum_{k=0}^{\infty}\frac{(At^2)^k}{(2k+1)!}\,.$$

(a) For $a > 0$ and $\omega = \sqrt{a}$, one has therefore

$$\exp(At) = \mathbb{1}\cosh(\sqrt{a}t) + A\frac{\sinh(\sqrt{a}t)}{\sqrt{a}} = \begin{pmatrix} \cosh(\omega t) & \sinh(\omega t)/\omega \\ \omega\sinh(\omega t) & \cosh(\omega t) \end{pmatrix}.$$

(b) For $a = 0$, one has $A^2 = 0$, hence $\exp(At) = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$.
(c) For $a < 0$, it follows with $\omega = \sqrt[2]{-a}$ that

$$\exp(At) = \mathbb{1}\cos(\omega t) + A\frac{\sin(\omega t)}{\omega} = \begin{pmatrix} \cos(\omega t) & \sin(\omega t)/\omega \\ \omega\sin(\omega t) & \cos(\omega t) \end{pmatrix}. \qquad \square$$

H.6 Chapter 6, Hamiltonian Equations and Symplectic Group

Exercise 6.23 on page 110 (Symplectic Algebra): u is infinitesimally symplectic. For the matrix U representing u with respect to a basis in which the symplectic bilinear form ω is represented by $\mathbb{J} = \left(\begin{smallmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{smallmatrix}\right)$, one has therefore $\mathbb{J}U + U^\top\mathbb{J} = 0$. The eigenvalues are the zeros of the characteristic polynomial

$$\begin{aligned} p_U(\lambda) &= \det(\lambda\mathbb{1} - U) = \det\big(\mathbb{J}(\lambda\mathbb{1} - U)\big) = \det(\lambda\mathbb{J} - \mathbb{J}U) = \det\big(\lambda\mathbb{J} + U^\top\mathbb{J}\big) \\ &= \det\big((\lambda\mathbb{1} + U^\top)\mathbb{J}\big) = \det\big(\lambda\mathbb{1} + U^\top\big) = \det(\lambda\mathbb{1} + U) = \det(-\lambda\mathbb{1} - U). \end{aligned}$$

So the characteristic polynomial is even, hence $-\lambda$ is an eigenvalue if λ is. As U has only real entries, $\overline{\lambda}$ will be an eigenvalue if λ is. Multiplicities of eigenvalues that are related in this manner coincide.

The even multiplicity of 0 results from the fact that the characteristic polynomial is even. $\square$

Exercise 6.26 on page 111 (Symplectic Matrices):

(a) $u^\top\mathbb{J} + \mathbb{J}u = \begin{pmatrix} A^\top & C^\top \\ B^\top & D^\top \end{pmatrix}\left(\begin{smallmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = \begin{pmatrix} C^\top - C & -A^\top - D \\ D^\top + A & -B^\top + B \end{pmatrix}$.
(b) This follows from the fact that the following expression is 0:

$$\begin{aligned} a^\top\mathbb{J}\,a - \mathbb{J} &= \begin{pmatrix} A^\top & C^\top \\ B^\top & D^\top \end{pmatrix}\begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} - \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \\ &= \begin{pmatrix} C^\top & -A^\top \\ D^\top & -B^\top \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} - \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} = \begin{pmatrix} C^\top A - A^\top C & C^\top B - A^\top D + \mathbb{1} \\ D^\top A - B^\top C - \mathbb{1} & D^\top B - B^\top D \end{pmatrix}. \end{aligned}$$

(c) $\mathrm{SL}(2,\mathbb{R}) = \{u \in \mathrm{Mat}(2,\mathbb{R}) \mid \det(u) = 1\}$ and the condition from (b) prove the first claim. The three-dimensional solid torus is $S^1 \times B$. As

$$\mathrm{SO}(2) = \left\{ \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \;\middle|\; \varphi \in S^1 \right\},$$

the S^1-coordinate is already identified. The positive symmetric matrices with determinant 1 can be written as $\left(\begin{smallmatrix} A & B \\ B & C \end{smallmatrix}\right)$ with $A > 0$, $B \in \mathbb{R}$ and $C = \frac{1+B^2}{A}$. The *Cayley transform* of the Riemann sphere

$$\mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\} \quad , \quad z \mapsto \tfrac{z-\imath}{z+\imath}$$

maps the upper half plane $\{z = B + \imath A \in \mathbb{C} \mid B \in \mathbb{R},\ A > 0\}$ onto the open unit disc $B = \{w \in \mathbb{C} \mid |w| < 1\}$ diffeomorphically, with its inverse being $w \mapsto \imath\,\frac{1+w}{1-w}$.

(d) Since $\left(\begin{smallmatrix} a & b \\ b & -a \end{smallmatrix}\right)^{2n} = (a^2+b^2)^n\,\mathbb{1}$, it follows

$$\exp\left(\begin{smallmatrix} a & b \\ b & -a \end{smallmatrix}\right) = \cosh\big(\sqrt{a^2+b^2}\big)\,\mathbb{1} + \frac{\sinh\big(\sqrt{a^2+b^2}\big)}{\sqrt{a^2+b^2}}\left(\begin{smallmatrix} a & b \\ b & -a \end{smallmatrix}\right),$$

if $a^2 + b^2 > 0$. Therefore, $\mathrm{tr}(M) = 2\cos(\varphi)\cosh\big(\sqrt{a^2+b^2}\big)$. The formula

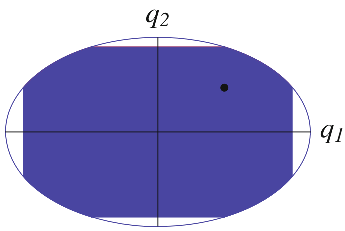
$$\tfrac{1}{2}\left(\mathrm{tr}(M) \pm \sqrt{\mathrm{tr}(M)^2 - 4\det(M)}\right)$$

for the eigenvalues of a matrix $M \in \mathrm{Mat}(2, \mathbb{R})$ shows that the eigenvalues of the symplectic matrix M coincide exactly when $|\mathrm{tr}(M)| = 2$. □

Exercise 6.34 on page 115 (Lissajous Figures):

1. The frequency ratio is $\dfrac{\omega_2}{\omega_1} = \dfrac{\text{number of maxima of } q_2}{\text{number of maxima of } q_1}$.

2. $\mathcal{E} := \{Q \in \mathbb{R}^2 \mid H(0, Q) \le E\} = \{Q \in \mathbb{R}^2 \mid \omega_1 Q_1^2 + \omega_2 Q_2^2 \le 2E\}$ is the elliptic Hill domain, from which we choose the initial point q. Then the closure of the Lissajous figure with initial condition (p, q), $H(p, q) = E$ is the rectangle $R_p := \{Q \in \mathbb{R}^2 \mid |Q_k| \le \sqrt{p_k^2 + q_k^2}\}$. The union of these rectangles is the subset



$$\mathcal{E} \,\cap\, [-R_1, R_1] \times [-R_2, R_2],$$

of the ellipse, where $R_1 := \sqrt{(2E - \omega_2 q_2^2)/\omega_1}$ and $R_2 := \sqrt{(2E - \omega_1 q_1^2)/\omega_2}$. No point $q' \in \mathcal{E}$ outside this domain can be reached from q, and for $q \neq 0$, there do exist such points q'. So the analog of the Hopf-Rinow theorem does not apply to Hill domains. □

Exercise 6.38 on page 118 (Linking Number):

- With the frequencies $\omega_k = n_k \omega_0$, the motion of the harmonic oscillator is periodic with minimal period $T = \frac{2\pi}{\omega_0}$. The normal oscillations on the energy shell $\Sigma_E = H^{-1}(E),\ E > 0$ can therefore be written in the form

$$\tilde{c}_k : S^1 \to \Sigma_E \quad , \quad \tilde{c}_1(t) = \sqrt{\frac{2E}{\omega_1}}\begin{pmatrix} \cos(\omega_1 t) \\ 0 \\ \sin(\omega_1 t) \\ 0 \end{pmatrix} \quad , \quad \tilde{c}_2(t) = \sqrt{\frac{2E}{\omega_2}}\begin{pmatrix} 0 \\ \cos(\omega_2 t) \\ 0 \\ \sin(\omega_2 t) \end{pmatrix},$$

where we understand the identification $S^1 := \mathbb{R}/T\mathbb{Z}$. By the linear mapping

$$(p_1, p_2, q_1, q_2)^\top \longmapsto (2E)^{-1/2}\big(\sqrt{\omega_1}p_1, \sqrt{\omega_2}p_2, \sqrt{\omega_1}q_1, \sqrt{\omega_2}q_2\big)^\top,$$

$\Sigma_E \subset \mathbb{R}^4$ is diffeomorphic to $S^3 \subset \mathbb{R}^4$. Let the projections of the normal oscillations be denoted as $\hat{c}_k : S^1 \to S^3$. We project[30] them onto $\mathbb{R}^3$ by means of the stereographic projection from Example A.29.3. The image curves of the $\hat{c}_k$ are obtained as

$$c_1 : S^1 \to \mathbb{R}^3 \quad , \quad c_1(t) = 2\begin{pmatrix} \cos(\omega_1 t) \\ 0 \\ \sin(\omega_1 t) \end{pmatrix} \quad \text{, hence} \quad c_1'(t) = 2\omega_1 \begin{pmatrix} -\sin(\omega_1 t) \\ 0 \\ \cos(\omega_1 t) \end{pmatrix}$$

$$c_2 : S^1 \to \mathbb{R}^3 \cup \{\infty\}, c_2(t) = \frac{2}{1-\sin(\omega_2 t)}\begin{pmatrix} 0 \\ \cos(\omega_2 t) \\ 0 \end{pmatrix} \text{, hence}$$
$$c_2'(t) = \frac{2\omega_2}{1-\sin(\omega_1 t)}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Thus $\Delta c(t) = \|c_1(t_1) - c_2(t_2)\| = 2\sqrt{\frac{2}{1-\sin(\omega_2 t_2)}}$ and

$$G(t_1, t_2) = \frac{1}{\sqrt{2}}\begin{pmatrix} \cos(\omega_1 t_1)\sqrt{1-\sin(\omega_2 t_2)} \\ -\cos(\omega_2 t_2)/\sqrt{1-\sin(\omega_2 t_2)} \\ \sin(\omega_1 t_1)\sqrt{1-\sin(\omega_2 t_2)} \end{pmatrix}.$$

For the linking number (6.3.9) of the two curves, we obtain, in view of

$$\det\big(\mathrm{D}G(t)\big) = -\omega_1\omega_2\sqrt{1-\sin(\omega_2 t_2)}\,,$$

the result

$$\begin{aligned} LK(c_1, c_2) &= -\frac{\omega_1\omega_2}{4\pi\sqrt{8}}\int_0^T\int_0^T \sqrt{1-\sin(\omega_2 t_2)}\,\mathrm{d}t_1\,\mathrm{d}t_2 \\ &= -\frac{n_1\omega_2}{4\sqrt{2}}\int_0^T \sqrt{1-\sin(\omega_2 t_2)}\,\mathrm{d}t_2 = -n_1 n_2\,. \end{aligned}$$

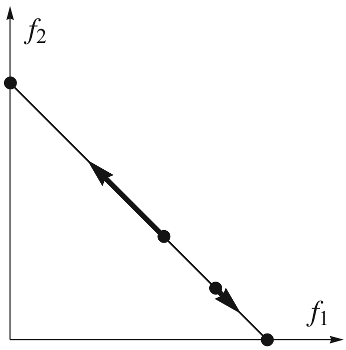
- For arbitrary pairs of distinct periodic trajectories in Σ_E, there exists a continuous homotopy of initial conditions connecting them with $\tilde{c}_1(0)$ and $\tilde{c}_2(0)$ respectively and avoiding the equality of the deformed orbits.

[30] It is not a problem for the integration that for $t = \left(\frac{\pi}{2} + 2\pi k\right)/\omega_2$, $k \in \mathbb{Z}$, the point $\hat{c}_2(t)$ is the north pole $(0, 0, 0, 1)^\top \in S^3$.

The figure shows the values of the mapping $\Sigma_E \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix}$, where the F_k are the constants of motion from (6.3.5). One obtains a straight segment because

$$F_k \geq 0 \quad \text{and} \quad F_1 + F_2 = H.$$

The effect of the homotopy on the values of F is indicated by arrows.



- In the mapping on page 116, the projected normal oscillation c_1 corresponds to the horizontal circle, and c_2 corresponds to the vertical straight line. This relation is not changed when the frequencies are different ($n_1 \neq n_2$). □

Exercise 6.40 on page 121 (Dispersion Relation):

(a) If in[31] the ansatz $q_\ell^{(a)}(t) := c^{(a)} \exp(2\pi\imath k\ell/n + \imath\omega_k t)$, the ratio of the constants is $\lambda := c^{(2)}/c^{(1)}$, then the equations for the change of momenta $\dot{p}_\ell^{(a)} = -\omega_k^2 m^{(a)} q_\ell^{(a)}$ yield the coupled system

$$\omega_k^2 m^{(1)} = c[2 - \lambda(e^{2\pi\imath k/n} + 1)] \quad , \quad \omega_k^2 m^{(2)} = c[2 - \lambda^{-1}(e^{-2\pi\imath k/n} + 1)] \,.$$

The quadratic equation for the squares of the frequencies ω_k^2 is therefore

$$\left(\omega_k^2 - \frac{2c}{m^{(1)}}\right)\left(\omega_k^2 - \frac{2c}{m^{(2)}}\right) - 2c^2 \frac{1 + \cos\left(\frac{2\pi k}{n}\right)}{m^{(1)} m^{(2)}} = 0 \,.$$

Its solutions are the two branches of the dispersion relation.

(b) From the Equation (6.3.12) for the frequencies ω_k, namely

$$\omega_k^2 = \frac{2}{m} \sum_{r \in \mathcal{L}} c_r \big(1 - \cos(2\pi k\, r/n)\big) \qquad (k \in \mathcal{L}),$$

the equation from which the coupling constants c_r are determined is obtained by Fourier transformation with respect to the group $\mathcal{L} = \mathbb{Z}/n\mathbb{Z}$: Only the sums $c_r + c_{-r}$ enter into the interaction, and they are proportional to the Fourier coefficients $\hat{\omega}_r^2 := \sum_{k \in \mathcal{L}} \omega_k^2 \exp(2\pi\imath k\, r/n)$. □

Exercise 6.42 on page 124 (Upper Half Plane and Möbius Transformations):

(a) We need to check $\Im\big(\widehat{M} z\big) > 0$.

[31] The end points of this segment are the values of F for the normal oscillations.

$$\Im(\widehat{M}z) = \tfrac{1}{2\imath}\left(\tfrac{az+b}{cz+d} - \overline{\tfrac{az+b}{cz+d}}\right) = \tfrac{1}{2\imath}\left(\tfrac{(az+b)(c\bar{z}+d)-(cz+d)(a\bar{z}+b)}{|cz+d|^2}\right)$$
$$= \frac{ad\,\Im z - bc\,\Im z}{|cz+d|^2} \overset{\det M=1}{=} \frac{\Im z}{|cz+d|^2} > 0\,.$$

(b) Let $M, M' \in \mathrm{SL}(2,\mathbb{R})$ and $z \in \mathbb{H}$. Then $M^{-1} = \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$ and $M^{-1}M' = \left(\begin{smallmatrix} a'd-bc' & b'd-bd' \\ ac'-a'c & ad'-b'c \end{smallmatrix}\right)$.

$$\widehat{M^{-1}}\widehat{M'}z = \widehat{M^{-1}}\frac{a'z+b'}{c'z+d'} = \frac{d\frac{a'z+b'}{c'z+d'} - b}{-c\frac{a'z+b'}{c'z+d'} + a} = \frac{d(a'z+b') - b(c'z+d')}{-c(a'z+b') + a(c'z+d')}$$
$$= \frac{(a'd-bc')z + b'd - bd'}{(ac'-a'c)z + ad' - cb'} = \widehat{M^{-1}M'}z\,,$$

so indeed we have a group action.

(c) Set up the equation for fixed points:

$$z = \widehat{M}z = \frac{az+b}{cz+d} \iff (cz+d)z = az+b \iff cz^2 + (d-a)z - b = 0.$$

This quadratic equation has discriminant

$$(d-a)^2 + 4bc = (\mathrm{tr}\,M)^2 - 4\,.$$

- For an elliptic matrix, the discriminant is negative, hence there exists exactly one fixed point in $\mathbb{H}$.
- For a parabolic matrix, the discriminant vanishes, hence there exist a (double) fixed point in $\mathbb{R} \cup \{\infty\}$; the fixed point is ∞ when $c = 0$.
- For a hyperbolic matrix, we obtain two fixed points in $\mathbb{R} \cup \{\infty\}$, because the discriminant is positive. When $c = 0$, one of them is ∞.

(d) The Möbius transformations with $M_1 := \left(\begin{smallmatrix} a & b \\ 0 & a^{-1} \end{smallmatrix}\right)$ and $M_2 := \left(\begin{smallmatrix} 0 & c \\ -c^{-1} & 0 \end{smallmatrix}\right)$ are

$$\widehat{M}_1 : z \mapsto a^2 z + ab \quad \text{and} \quad \widehat{M}_2 : z \mapsto c^2 z^{-1} = \frac{c^2\bar{z}}{|z|^2} = c^2\frac{x - \imath y}{x^2+y^2}\,.$$

Their derivatives at $z = x + \imath y \in \mathbb{H}$ are

$$\mathrm{D}\widehat{M}_1(x,y) = \begin{pmatrix} a^2 & 0 \\ 0 & a^2 \end{pmatrix} \quad \text{and} \quad \mathrm{D}\widehat{M}_2(x,y) = \frac{c^2}{(x^2+y^2)^2}\begin{pmatrix} y^2-x^2 & -2xy \\ 2xy & y^2-x^2 \end{pmatrix}.$$

We transport the metric: $\widehat{M}_1^* g = (a^2 y)^{-2}(a^4 \mathrm{d}x \otimes \mathrm{d}x + a^4 \mathrm{d}y \otimes \mathrm{d}y) = g$, and with $v = (v_x, v_y) \in T_z\mathbb{H}$:

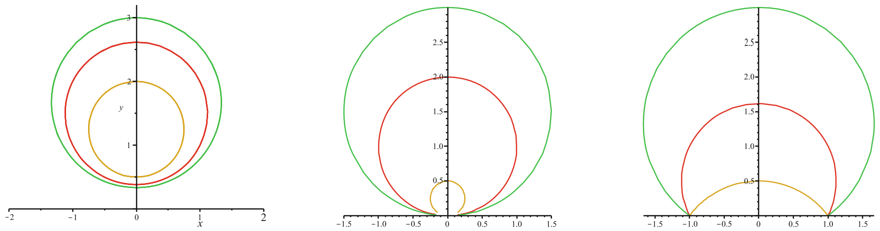


Figure H.2 Orbits of the subgroups from 6.42 (f) through the points $1+\imath$ (red), $3\imath$ (green) and $\imath/2$ (yellow). Left: elliptic, middle: parabolic, right: hyperbolic (Image: courtesy of Christoph Schumacher)

$$\begin{aligned}\big(\widehat{M}_2^* g\big)(z)(v,v) &= g\big(\widehat{M}_2 z\big)\ \big(\mathrm{D}\widehat{M}_2 v, \mathrm{D}\widehat{M}_2 v\big)\\ &= \frac{(x^2+y^2)^2}{(c^2y)^2}\left\| \frac{c^2}{(x^2+y^2)^2}\left(\begin{smallmatrix}(y^2-x^2)v_x-2xyv_y\\ 2xyv_x+(y^2-x^2)v_y\end{smallmatrix}\right)\right\|^2\\ &= \frac{((y^2-x^2)v_x-2xyv_y)^2+(2xyv_x+(y^2-x^2)v_y)^2}{y^2(x^2+y^2)^2}\\ &= \frac{(y^2-x^2)^2(v_x^2+v_y^2)+4x^2y^2(v_x^2+v_y^2)}{y^2(x^2+y^2)^2} = \frac{\|v\|^2}{y^2} = g(z)(v,v)\,.\end{aligned}$$

The matrices that we have checked generate $\mathrm{SL}(2,\mathbb{R})$. Indeed, if $M = \left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{R})$ is not of the form M_1, then $c \neq 0$. In this case, if M is not of the form M_2, then either $a \neq 0$ or $d \neq 0$. By conjugating, one obtains $\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right) M \left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right) = \left(\begin{smallmatrix}-d&c\\b&-a\end{smallmatrix}\right)$. Following with a multiplication by a matrix of type M_1, one can set the top right entry to 0. Conjugating again with $\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$ produces a matrix of type M_1.

Thus g is invariant under all Möbius transformations.

(e) follows from Exercise 6.26 (a).

(f) The cross ratio $[z_1, z_2, z_3, z_4] := \frac{z_1-z_2}{z_1-z_4}\cdot\frac{z_3-z_4}{z_3-z_2}$ is real exactly if the four points $z_1, z_2, z_3, z_4 \in \mathbb{C}$ lie on a circle.

- $m = \left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$: $\exp\big(t\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)\big) = \left(\begin{smallmatrix}\cos t&\sin t\\-\sin t&\cos t\end{smallmatrix}\right)$ is elliptic (except when $t \in \pi\mathbb{Z}$) with fixed point $\imath \in \mathbb{H}$. The orbits are circles in $\mathbb{H}$, see Figure H.2, left.
- $m = \left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right)$: $\exp\big(t\left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right)\big) = \left(\begin{smallmatrix}1&0\\t&1\end{smallmatrix}\right)$ is parabolic (only for $t = 0$ is it equal to $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$) with fixed point $0 \in \partial\mathbb{H}$. The closure of the orbit is a circle that is tangential to $\partial\mathbb{H}$ in the origin, see Figure H.2, middle.
- $m = \left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$: $\exp\big(t\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)\big) = \left(\begin{smallmatrix}\cosh t&\sinh t\\\sinh t&\cosh t\end{smallmatrix}\right)$ is hyperbolic (except when $t = 0$) with fixed points $-1, 1 \in \partial\mathbb{H}$. The orbits are circular segments that are delimited by fixed points, see Figure H.2, right. □

Exercise 6.45 on page 127 (Symplectic Mappings and Subspaces):

(a) Let $v, w \in E\backslash\{0\}$. The antisymmetric bilinear form $\omega : E \times E \to \mathbb{R}$ of the $2n$-dimensional vector space E is non-degenerate. So by the linear Darboux theorem

(Theorem 6.13.2), there exist two bases $d_1, \ldots, d_{2n}$ and $e_1, \ldots, e_{2n}$ of E with $d_1 = v,\ e_1 = w$ and

$$\omega(d_k, d_\ell) = \omega(e_k, e_\ell) = \delta_{k+n,\ell} \qquad (1 \le k \le \ell \le 2n). \tag{H.6.1}$$

The linear mapping $f : E \to E$ defined by the change of bases $f(d_k) := e_k$ $(k = 1, \ldots, 2n)$ is in $\mathrm{Sp}(E, \omega)$.

(b) Let $E := \mathbb{R}^4$, with canonical basis $e_1, \ldots, e_4$ and the symplectic form ω that is determined by $\omega(e_k, e_\ell) = \delta_{k+2,\ell}$ $(1 \le k \le \ell \le 4)$. Then $F := \mathrm{span}(e_1, e_2)$ is Lagrangian and $F' := \mathrm{span}(e_1, e_3)$ is symplectic. Therefore, there cannot exist $f \in \mathrm{Sp}(E, \omega)$ with $f(F) = F'$.

(c) Let $F, F' \subseteq E$ be symplectic and of the same dimension $2m$. The proof of Theorem 6.13.2 provides the existence of bases $d_1, \ldots, d_{2n}$ and $e_1, \ldots, e_{2n}$ of E with (H.6.1) and $d_1, \ldots, d_{2m} \in F, e_1, \ldots, e_{2m} \in F'$. Therefore, the mapping $f \in \mathrm{Sp}(E, \omega)$ defined by $f(d_k) := e_k$ maps F onto F'. □

Exercise 6.47 on page 127 (Dimension Formula):
If $f_1, \ldots, f_m$ is a basis of $F \subseteq E$, then the linear forms $f_1^*, \ldots, f_m^* \in E^*$ given by $f_k^*(e) := \omega(f_k, e)$ for $e \in E$ are linearly independent, because ω is non-degenerate. Therefore,

$$F^\perp = \{e \in E \mid f_1^*(e) = \ldots = f_m^*(e) = 0\}$$

has dimension $\dim(F^\perp) = \dim(E) - \dim(F)$. □

Exercise 6.53 on page 131 ($\mathrm{SO}(\mathbf{3}) \cong \mathbb{R}\mathrm{P}(\mathbf{3})$):

- The Rodrigues parametrization $A \in C^\infty\big(B_\pi^3, \mathrm{SO}(3)\big)$ from (E.3.3) is surjective, and its restriction to the *open* ball of radius π is a diffeomorphism onto its image. This can be seen, for example, by checking that for $\|x\| \in (0, \pi)$, the inversion of $A(x)$ is given by $\frac{r}{2\sin(r)}(A(x) - A(x)^\top)$ with $r := \arccos\left(\frac{\mathrm{tr}(A(x))-1}{2}\right)$.
  At $x \in \partial B_\pi^3$, i.e., $\|x\| = \pi$, the mapping A is still a local diffeomorphism. If we identify the antipodes of the 2-sphere ∂B_π^3 (denoting the identification relation $\sim$), then A becomes injective.
- On the other hand, $B_\pi^3/\sim$ is diffeomorphic to $\mathbb{R}P(3)$. This is because under a stereographic projection $S^3\backslash\{(0, 0, -1)^\top\} \to \mathbb{R}^3$ (scaled by $\pi/2$), the manifold with boundary B_π^3 is diffeomorphic to the northern hemisphere $\{x \in S^3 \mid x_3 \ge 0\}$, see for instance A.29.3. Moreover, every straight line through the origin in $\mathbb{R}^3$ (i.e., every element of $\mathbb{R}P(3)$) has exactly one intersection in the northern hemisphere, except for those that hit the equator $\{x \in S^3 \mid x_3 = 0\}$ in antipodes. □

Exercise 6.59 on page 134 (Maslov Index):

1. For $u\in\Lambda_k(m)=\{u \in \Lambda(m) \mid \dim(u \cap v)=k\}$ with $v = \mathbb{R}_p^m \times \{0\} \subset \mathbb{R}_p^m \times \mathbb{R}_q^m$, we set $u_v := u \cap v$. For $w \in \mathrm{Gr}(v, k)$, the fiber $\pi_k^{-1}(w)$ consists of those $u \in$

$\Lambda_k(m)$ for which $u_v = w$. It can be parametrized by the subspace $\mathrm{Sym}(w^\perp)$ of self-adjoint mappings from

$$w^\perp := \{x \in \mathbb{R}^m \mid \forall\, y \in w : \ \langle x, y\rangle = 0\}$$

into itself, because their graph is Lagrangian, and the symplectic vector space $(\mathbb{R}^m \times \mathbb{R}^m, \omega_0)$ is the direct sum of its symplectic subspaces $w \times w$ and $w^\perp \times w^\perp$.
On the other hand, $\dim\big(\mathrm{Sym}\big(w^\perp\big)\big) = \binom{m-k+1}{2}$, because $\dim\big(w^\perp\big) = m - k$.
Locally in a neighborhood of $w \in \mathrm{Gr}(v, k)$, the bundle $\pi_k : \Lambda_k(m) \to \mathrm{Gr}(v, k)$ can be trivialized by means of orthogonal projections.
As the dimension of the total space of a bundle is the sum of the dimensions of the base and of a typical fiber, it follows from $\dim\big(\mathrm{Gr}(v, k)\big) = k(m - k)$ (Theorem 6.51) that

$$\dim\big(\Lambda_k(m)\big) = k(m-k) + \tfrac{1}{2}(m-k+1)(m-k) = \binom{m+1}{2} - \binom{k+1}{2}.$$

2. The parametrization of $\Lambda_0(m)$ by $\mathrm{Sym}(\mathbb{R}^m)$, given in part 1, shows that $\Lambda_0(m) \subset \Lambda(m)$ is open. If $u \in \Lambda_k(m)$ for $k \geq 2$, then in $u_v \in \mathrm{Gr}(v, k)$, one can choose an orthogonal decomposition $u_v = a \oplus b$ into a one-dimensional subspace a and $(k-1)$-dimensional subspace b. This way, u_v is approximated by a sequence of Lagrangian subspaces

$$(\{0\} \times a) \oplus \mathrm{graph}(n\, 1\!\!1_b) \qquad (n \in \mathbb{N}),$$

where we have used the identity map $1\!\!1_b : b \to b$. □

Exercise 6.61 on page 134 (Range of the Maslov Index):
• For $m = 1$ and $I \in \mathbb{Z}$, the mapping

$$c : S^1 \to \Lambda(1) \quad , \quad z \mapsto \mathrm{span}_{\mathbb{R}}\big(z^{I/2}\big) \ \subset \mathbb{C} \cong \mathbb{R}^2$$

has the desired Maslov index I. We have used the identifications

$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\} \quad \text{and} \quad \Lambda(1) = \{\mathrm{span}_{\mathbb{R}}(z) \mid z \in S^1\}.$$

Note that $\mathrm{span}_{\mathbb{R}}\big(z^{I/2}\big)$ is well-defined even if I is odd.
• For $m > 1$, one can imbed $\Lambda(1)$ into $\Lambda(m)$, say by

$$\Lambda(1) \to \Lambda(m) \quad , \quad u \mapsto u \oplus \mathbb{R}^{m-1} \times \{0\}.$$ □

Exercise 6.62 on page 135 (Harmonic Oscillator):
For $f = (f_1, f_2)$ with $f_i > 0$ and $k_i = 2 f_i/\omega_i$, any such a closed solution curve $\tilde{c} : S^1 \to \mathbb{T}^2$ on the Lagrange torus $\mathbb{T}^2 := F^{-1}(f)$ can be represented as

$$\tilde{c}(t) = \big(k_1 \sin(\omega_1(t-t_1)), k_2 \sin(\omega_2(t-t_2)), k_1 \cos(\omega_1(t-t_1)), k_2 \cos(\omega_2(t-t_2))\big),$$

with t_1, t_2 chosen appropriately. The invariant torus $\mathbb{T}^2 \subset \mathbb{R}^2_p \times \mathbb{R}^2_q$ in phase space projects onto the rectangle $[-k_1, k_1] \times [-k_2, k_2] \subset \mathbb{R}^2_q$ in configuration space. In order to avoid double imaginary eigenvalues below, we choose a curve that does not hit a vertex of that rectangle. Along the four intersecting circles

$$\big\{(0, k_2 \sin\varphi, \pm k_1, k_2 \cos\varphi) \mid \varphi \in [0, 2\pi]\big\} \subset \mathbb{T}^2$$

and $\{(k_1 \sin\varphi, 0, k_1 \cos\varphi, \pm k_2) \mid \varphi \in [0, 2\pi]\} \subset \mathbb{T}^2$, this projection is vertical.

Within a period interval $[0, T]$, with $T = \frac{6\pi}{\omega_1} = \frac{10\pi}{\omega_2}$, these curves will be hit by $\tilde{c}$ six or ten times respectively, in total 16 times.

From the projection $t \mapsto \big(k_1 \cos(\omega_1(t - t_1)),\ k_2 \cos(\omega_2(t - t_2))\big)$ of $\tilde{c}$ onto the configuration space, one can see that the 1-component has six, and the 2-component ten extrema. As the orientation sign $\left(\frac{\lambda'}{\imath\lambda}\right)$ for the unique imaginary eigenvalue λ of a unitary representation U of c equals one for these 16 times, the signs add up to the Maslov index $\deg(\mathrm{MA}_2 \circ \tilde{c}) = 16$ of the closed curve. □

H.7 Chapter 7, Stability Theory

Exercise 7.4 on page 140 (Strong Stability):

To begin with,

$$X_H : \mathbb{R}_t \times \mathbb{R}^2_x \to \mathbb{R}^2 \quad , \quad X_H(t, x) = \begin{pmatrix} 0 & -f(t) \\ 1 & 0 \end{pmatrix} x$$

is the time dependent Hamiltonian vector field for H. The differential equation is solvable for all times by Theorem 4.14, because X_H is linear in x and thus satisfies a Lipschitz condition.

(a) Counterexample to the group property: For the 2-periodic characteristic function $f := 1\!\!1_{[0,1]+2\mathbb{Z}}$, the existence and uniqueness of a solution is still guaranteed, because the Lipschitz condition is only required with respect to x (see Remark 3.24.3). We can solve the differential equation piecewise: $x(t) = M_k(\{t\})x(\lfloor t \rfloor)$ with $M_0(t) := \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}$, $M_1(t) := \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$ and $k := 0$ for even $\lfloor t \rfloor$, $k := 1$ for odd $\lfloor t \rfloor$. Then $M_0^2(1) \neq M_0(1)M_1(1)$.

(b) For $s = 0$, one has $\Phi_{T+s} = \Phi_0 \circ \Phi_T$, because $\Phi_0 = \mathrm{Id}$. Furthermore,

$$\frac{\mathrm{d}}{\mathrm{d}s}\Phi_{T+s}(x) = X_H\big(T + s, \Phi_{T+s}(x)\big) = \begin{pmatrix} 0 & -f(s) \\ 1 & 0 \end{pmatrix} \Phi_{T+s}(x)\,,$$

because $f(T + s) = f(s)$ and

$$\frac{\mathrm{d}}{\mathrm{d}s}\Phi_s \circ \Phi_T(x) = X_H\big(s, \Phi_s \circ \Phi_T(x)\big) = \begin{pmatrix} 0 & -f(s) \\ 1 & 0 \end{pmatrix} \Phi_s \circ \Phi_T(x)\,.$$

Thus the functions $s \mapsto \Phi_s \circ \Phi_T(x)$ and $s \mapsto \Phi_{T+s}(x)$ both satisfy the same uniquely solvable initial value problem, so they must coincide.

(c) The mapping $\Phi_t\colon \mathbb{R}^2 \to \mathbb{R}^2$ is linear for all $t \in \mathbb{R}$, because for $t = 0$, one has $\Phi_0 = \mathrm{Id}$, which is obviously linear, and furthermore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t(\lambda x + y) = X_H\big(t, \Phi_t(\lambda x + y)\big) = \left(\begin{smallmatrix} 0 & -f(t) \\ 1 & 0 \end{smallmatrix}\right) \Phi_t(\lambda x + y)$$

and $\frac{\mathrm{d}}{\mathrm{d}t}\big(\lambda\Phi_t(x) + \Phi_t(y)\big) =$

$$\lambda X_H\big(t, \Phi_t(x)\big) + X_H\big(t, \Phi_t(y)\big) = \left(\begin{smallmatrix} 0 & -f(t) \\ 1 & 0 \end{smallmatrix}\right) \big(\lambda\Phi_t(x) + \Phi_t(y)\big)\,.$$

So the functions satisfy the same initial value problem, hence coincide.
The Wronskian $t \mapsto \det(\Phi_t)$ has the derivative $\frac{\mathrm{d}}{\mathrm{d}t}\det(\Phi_t) =$

$$\mathrm{tr}\big(\det(\Phi_t)\Phi_t^{-1}\dot{\Phi}_t\big) = \det(\Phi_t)\mathrm{tr}\big(\Phi_t^{-1}X_H(t)\Phi_t\big) = \det(\Phi_t)\mathrm{tr}\big(X_H(t)\big) = 0\,.$$

From $\det(\Phi_0) = 1$, it follows in particular that $\det(A) = 1$. With $\mathbb{J} = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ we conclude

$$\mathrm{tr}\big(X_H(t)\big) = \mathrm{tr}\big(\mathbb{J}B(t)\big) = \mathrm{tr}\big(B^\top(t)\mathbb{J}^\top\big) = -\mathrm{tr}\big(B(t)\mathbb{J}\big) = -\mathrm{tr}\big(\mathbb{J}B(t)\big) = 0.$$

(d) If the zero solution of the Hamiltonian flow is Lyapunov-stable, then 0 is also a Lyapunov-stable fixed point of the mapping A, because $\Psi = \Phi\restriction_{\mathbb{Z}\times\mathbb{R}^2}$. To see the converse, let $\kappa := (\sup\{\|\Phi_s\| \mid s \in [0, T]\})^{-1}$, and let U be a neighborhood of $0 \in \mathbb{R}^2$. We choose $\varepsilon > 0$ such that $B_\varepsilon(0) \subseteq U$, and from the Lyapunov-stability of Ψ, we choose $\delta > 0$ such that $\Psi(n, x_0) \in B_{\varepsilon\kappa}(0)$ for all $n \in \mathbb{N}_0$ and for all initial conditions $x_0 \in B_\delta(0)$. For $t = nT + s$ with $n \in \mathbb{N}_0$ and $s \in [0, T[$, we then obtain

$$\|\Phi_t(x)\| = \|\Phi_s(\Psi(n, x))\| \le \|\Phi_s\|\|\Psi(n, x)\| \le \kappa^{-1}\varepsilon\kappa = \varepsilon\,.$$

Hence Φ is Lyapunov-stable.

(e) Ψ is Lyapunov-stable if $|\mathrm{tr}(A)| < 2$, because the characteristic polynomial of A is

$$\begin{aligned}\chi_A(\lambda) &= \lambda^2 - \lambda\,\mathrm{tr}(A) + \det(A)\\ &\overset{\det A=1}{=} \left(\lambda - \tfrac{1}{2}\mathrm{tr}A + \imath\sqrt{1 - (\tfrac{1}{2}\mathrm{tr}A)^2}\right)\left(\lambda - \tfrac{1}{2}\mathrm{tr}A - \imath\sqrt{1 - (\tfrac{1}{2}\mathrm{tr}A)^2}\right).\end{aligned}$$

The eigenvalues of A have absolute value 1, as can be easily checked. So A is a rotation, hence Lyapunov-stable.
We let $B(t) := \left(\begin{smallmatrix} 1 & 0 \\ 0 & f(t) \end{smallmatrix}\right)$ $(t \in \mathbb{R})$ and $U := \{\tilde{H} \mid \|\tilde{H} - H\| < \delta\}$, with $\delta > 0$ yet to be determined. Now let

$$K := \sup\left\{\|\Phi_s^{-1}\|, \|\tilde{\Phi}_s\| \mid s \in [0,T],\, \tilde{\Phi} \text{ generated by } \tilde{H} \in U\right\}.$$

We conclude $K < \infty$ as follows:

$$\|\tilde{\Phi}_t\| = \left\|\tilde{\Phi}_0 + \int_0^t \mathbb{J}\,\tilde{B}(s)\tilde{\Phi}(s)\,\mathrm{d}s\right\| \le \|\tilde{\Phi}_0\| + \int_0^t \|\tilde{B}(s)\|\|\tilde{\Phi}(s)\|\,\mathrm{d}s$$
$$\le 1 + \int_0^t (\|B(s)\| + \delta)\|\tilde{\Phi}(s)\|\,\mathrm{d}s\,,$$

and Gronwall's lemma now tells us that

$$\|\tilde{\Phi}_t\| \le \exp\left(\int_0^t (\|B(s)\| + \delta)\,\mathrm{d}s\right) \le \exp\left(\int_0^T (\|B(s)\| + \delta)\,\mathrm{d}s\right).$$

We compare the two time evolutions:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_t^{-1}\tilde{\Phi}_t) = \Phi_t^{-1}\dot{\Phi}_t\Phi_t^{-1}\tilde{\Phi}_t + \Phi_t^{-1}\dot{\tilde{\Phi}}_t = \Phi_t^{-1}\mathbb{J}\left(B(t) - \tilde{B}(t)\right)\tilde{\Phi}_t$$

and obtain

$$\|\tilde{\Phi}_t - \Phi_t\| \;\le\; \|\Phi_t\|\|\Phi_t^{-1}\tilde{\Phi}_t - 1\!\!1\| \;=\; \|\Phi_t\|\left\|\int_0^t \frac{\mathrm{d}}{\mathrm{d}s}(\Phi_s^{-1}\tilde{\Phi}_s)\,\mathrm{d}s\right\|$$
$$\le\; \|\Phi_t\|\int_0^t \|\Phi_t^{-1}\|\;\|B(t) - \tilde{B}(t)\|\;\|\tilde{\Phi}_t\|\,\mathrm{d}s \;\le\; K^3 T\delta\,.$$

As the trace is continuous, all $\tilde{\Phi}$ are Lyapunov-stable, and the fixed point 0 is strongly stable. Matrices whose trace has absolute value less than 2 are also called elliptic.

(f) For $\varepsilon = 0$, the differential equation reduces to $\ddot{x}(t) = -\omega^2 x(t)$, and its flow is

$$\Phi_t = \begin{pmatrix} \cos(\omega t) & -\omega\sin(\omega t) \\ \omega^{-1}\sin(\omega t) & \cos(\omega t) \end{pmatrix}.$$

So $T = 2\pi$, and accordingly

$$A = \Phi_{2\pi} = \begin{pmatrix} \cos(2\pi\omega) & -\omega\sin(2\pi\omega) \\ \omega^{-1}\sin(2\pi\omega) & \cos(2\pi\omega) \end{pmatrix}$$

with $|\mathrm{tr}(A)| = 2|\cos(2\pi\omega)| < 2$, because $2\omega \notin \mathbb{Z}$. □

Exercise 7.8 on page 143 (Lyapunov Function):

(a) $\nabla V(x) = x$ and $f_1(x) = -\|x\|^2 x$, hence $\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) = -\|x(t)\|^4 < 0$ for $x \in \mathbb{R}^2 \setminus \{0\}$. So the origin is asymptotically stable.
(b) The origin is unstable, because $f_2 = -f_1$.
(c) The form $f_3(x) = (1 + x_1)\binom{-x_2}{x_1}$ of the vector field implies that $\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) = 0$; therefore the orbits are contained in circles around the origin. Each point on

the line $x_1 = -1$ corresponds to an orbit, likewise the circular segments that are obtained by intersecting these circles with the half planes $(-1, \infty) \times \mathbb{R}$ and $(-\infty, -1) \times \mathbb{R}$.

(d) $V : \mathbb{R}^3 \to \mathbb{R},\ x \mapsto \frac{1}{2}\|x\|^2$ is a Lyapunov function, with

$$\langle f_4(x), \nabla V(x)\rangle = -x_1^2\|x\|^2 - x_2^4 - x_3^6 < 0 \quad \text{for} \quad x \in \mathbb{R}^3 \setminus \{0\}\,.$$

Hence $0 \in \mathbb{R}^3$ is asymptotically stable.

(e) We have $A := \mathrm{D}f(0) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, hence $\exp(At) = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The origin of this linearized system is therefore Lyapunov-stable, but not asymptotically stable. In particular, the x_3-axis consists of fixed points. □

Exercise 7.21 on page 151 (Parametrized Periodic Orbits):
By assumption (7.3.1) and Theorem 3.48, the mapping $(t, m, p) \mapsto \Phi_t^{(p)}(m)$ is n times continuously differentiable.

For the parameter $p_0 \in P$, there exists by Theorem 7.17 a Poincaré map for $\Phi^{(p_0)}$ at the point $m_0 \in M$ of the $\Phi^{(p_0)}$-periodic orbit. From a transversality argument, the existence of a neighborhood $\tilde{P} \subseteq P$ of p_0 follows, with a hypersurface $S \subset M$ through m_0 that is transversal to all flows $\Phi^{(p)},\ \ p \in \tilde{P}$. So there also exist a neighborhood $U \subset S$ of m_0 that is open in S, and Poincaré times $\tilde{T} \in C^n(U \times \tilde{P}, \mathbb{R}^+)$ with Poincaré map $F^{(p)}(u) := \Phi^{(p)}\big(\tilde{T}(u, p), u\big) \in S$.

By hypothesis, $F^{(p_0)}(m_0) = t_0$; and as $\tilde{T} \in C^n(U \times \tilde{P}, \mathbb{R}^+)$, also $(u, p) \mapsto F^{(p)}(u)$ is n times continuously differentiable. Also by hypothesis, 1 is not an eigenvalue of $\mathrm{D}F^{(p_0)}(m_0)$. Thus we obtain from Theorem 7.12 a mapping $\widehat{m} \in C^n(\tilde{P}, S)$, with $\widehat{m}(p_0) = m_0$, for which $F^{(p)}\big(\widehat{m}(p)\big) = \widehat{m}(p)$; this mapping therefore parametrizes the fixed point.[32]

We may also assume that $\mathrm{D}F^{(p)}\big(\widehat{m}(p)\big)$ has no eigenvalue 1 for $p \in \tilde{P}$. The minimal period $t : \tilde{P} \to (0, \infty)$ is the mapping defined by $t(p) = \tilde{T}\big(\widehat{m}(p), p\big)$.

While there could be further periodic orbits in arbitrarily small neighborhoods of $\widehat{m}(p_0)$, these could not have minimal periods that are close to $t(p_0)$. □

H.8 Chapter 8, Variational Principles

Exercise 8.5 on page 158 (Legendre Transform):

(a) From $r > 1$, we infer that $H \in C^1(\mathbb{R}^d, \mathbb{R})$ and is strictly convex. For $q \in \mathbb{R}^d$, we want to find the vector $\hat{p} = \hat{p}(q) \in \mathbb{R}^d$ that satisfies

$$\nabla_p\big(\langle p, q\rangle - H(p)\big)\Big|_{p=\hat{p}} = 0\,.$$

[32] Here as well as in the sequel, it may be necessary to reduce domains like $\tilde{P}$.

Because $\nabla H(0) = 0$, we have $\hat{p}(0) = 0$. Otherwise, $q = \|\hat{p}\|^{r-2}\hat{p}$, hence $\|q\| = \|\hat{p}\|^{r-1}$, which can be written as $\|\hat{p}\| = \|q\|^{s-1}$ with $(r-1)(s-1) = 1$. Therefore,

$$\hat{p} = \frac{q}{\|\hat{p}\|^{r-2}} = \frac{q}{\|q\|^{(r-2)(s-1)}} \overset{rs=r+s}{=} \|q\|^{s-2} q \,.$$

This allows us to calculate H^*:

$$H^*(q) = \langle q, \|q\|^{s-2}q\rangle - H\big(\|q\|^{s-2}q\big) = \|q\|^s - \tfrac{1}{r}\|q\|^{(s-1)r} = \tfrac{1}{s}\|q\|^s \,.$$

(b) We proceed like in (a).

$$0 = \nabla_p\big(\langle p, q\rangle - H(p)\big)\Big|_{p=\hat{p}} = q - A\hat{p} - b \,,$$

hence $q = A\hat{p} + b$, or $\hat{p} = A^{-1}(q-b)$. Again we calculate H^*:

$$\begin{aligned} H^*(q) &= \langle q, A^{-1}(q-b)\rangle - H\big(A^{-1}(q-b)\big) \\ &= \langle q, A^{-1}(q-b)\rangle - \tfrac{1}{2}\langle A^{-1}(q-b), AA^{-1}(q-b)\rangle - \langle b, A^{-1}(q-b)\rangle - c \\ &= \langle q-b, A^{-1}(q-b)\rangle - \tfrac{1}{2}\langle q-b, A^{-1}(q-b)\rangle - c \\ &= \tfrac{1}{2}\langle q-b, A^{-1}(q-b)\rangle - c \,. \end{aligned}$$

□

Exercise 8.8 on page 160 (Legendre-Transformation):

(a) Polar coordinates: $x = r\cos\varphi$, $y = r\sin\varphi$, $v = \begin{pmatrix} v_r\cos\varphi - v_\varphi r\sin\varphi \\ v_r\sin\varphi + v_\varphi r\cos\varphi \end{pmatrix}$.

$$L\left(\begin{pmatrix} r \\ \varphi \end{pmatrix}, \begin{pmatrix} v_r \\ v_\varphi \end{pmatrix}\right) = \tilde{L}\left(\begin{pmatrix} x(r,\varphi) \\ y(r,\varphi) \end{pmatrix}, \begin{pmatrix} v_r\cos\varphi - v_\varphi r\sin\varphi \\ v_r\sin\varphi + v_\varphi r\cos\varphi \end{pmatrix}\right) = \tfrac{1}{2}(v_r^2 + r^2 v_\varphi^2) - U(r).$$

The Legendre transform of this quantity is

$$H\left(\begin{pmatrix} r \\ \varphi \end{pmatrix}, \begin{pmatrix} p_r \\ p_\varphi \end{pmatrix}\right) = \tfrac{1}{2}\big(p_r^2 + r^{-2}p_\varphi^2\big) + U(r)\,.$$

(b) Since $\nabla_v L(q, v) = mv + \frac{e}{c}A(q)$, it follows

$$\begin{aligned} &\sup\left\{\langle p, v\rangle - L(q,v) \mid v \in \mathbb{R}^2\right\} \\ &= \langle p, \tfrac{1}{m}(p - \tfrac{e}{c}A(q))\rangle - \tfrac{1}{2m}\|p - \tfrac{e}{c}A(q)\|^2 + e\phi(q) - \tfrac{e}{cm}\langle p - \tfrac{e}{c}A(q), A(q)\rangle \\ &= \tfrac{1}{2m}\|p - \tfrac{e}{c}A(q)\|^2 + e\phi(q)\,. \end{aligned}$$

□

Exercise 8.12 on page 164 (Bead on a Wire):

(a) At time $t \in \mathbb{R}$, the parabolic wire is parametrized by

$$w \longmapsto \tilde{q}(t,w) = \begin{pmatrix} w\cos(\omega t) \\ w\sin(\omega t) \\ \frac{\alpha^2}{2} w^2 \end{pmatrix}.$$

If the parameter has the value $w(t)$, then, for $q(t) := \tilde{q}\big(t, w(t)\big)$, one obtains

$$\dot{q}(t) = \begin{pmatrix} \dot{w}(t)\cos(\omega t) - \omega w(t)\sin(\omega t) \\ \dot{w}(t)\sin(\omega t) + \omega w(t)\cos(\omega t) \\ \alpha^2 w(t)\dot{w}(t) \end{pmatrix},$$

hence $\|\dot{q}\|^2 = \dot{w}^2 + w^2\omega^2 + \alpha^4 w^2 \dot{w}^2$. Therefore, the Lagrangian (without explicit time dependence) has the form

$$L(w,\dot{w}) = \tfrac{m}{2}\big(1+\alpha^4 w^2\big)\dot{w}^2 \; + \; \tfrac{m}{2}\big(\omega^2 - g\alpha^2\big)w^2 .$$

The momentum conjugate to w is $p_w = \mathrm{D}_2 L(w,\dot{w}) = m(1+\alpha^4 w^2)\dot{w}$. Therefore, letting $c := g\alpha^2 - \omega^2$, the Hamiltonian equals

$$H(p_w, w) = \frac{p_w^2}{2m(1+\alpha^4 w^2)} + V(w) \quad \text{with} \quad V(w) = \tfrac{m}{2} c\, w^2 .$$

(b) The linearization $A := \mathrm{D}X_H(0)$ of the Hamiltonian vector field $X_H = \left(\begin{smallmatrix} -\mathrm{D}_2 H \\ \mathrm{D}_1 H \end{smallmatrix}\right)$ in the origin is of the form $A = \left(\begin{smallmatrix} 0 & -mc \\ \frac{1}{m} & 0 \end{smallmatrix}\right)$. Therefore $\det(A) = c$. Hence for $c > 0$ (i.e., a slow rotation), the equilibrium $(p_w, w) = (0,0)$ is an elliptic fixed point and the motion is Lyapunov-stable.

(c) For slow rotation ($c > 0$), the period of the libration motion in dependence on the energy E is determined by

$$T(E) = 4\int_0^{w_{\max}} \frac{\mathrm{d}w}{\dot{w}(E,w)} \quad \text{with } \dot{w}(E,w) = \sqrt{\frac{2(E-\beta^2 w^2)}{m(1+\alpha^4 w^2)}} \text{ and } w_{\max} = \frac{\sqrt{E}}{\beta}.$$

One obtains

$$T(E) = \frac{\sqrt{8m}}{\beta}\int_0^1 \sqrt{\frac{1+\frac{E\alpha^4}{\beta^2}y^2}{1-y^2}}\,\mathrm{d}y = \frac{\sqrt{8m}}{\beta}\int_0^{\pi/2}\sqrt{1+\frac{E\alpha^4}{\beta^2}\sin^2\theta}\,\mathrm{d}\theta . \quad \square$$

Exercise 8.20 on page 169 (Example for Non-Minimality of the Action Functional):

(a) The extremals satisfy the Euler-Lagrange equation (8.3.4), hence $\ddot{q} = -(0, q_2)^\top$. The family of solutions with initial condition $q(0) = 0$ is of the form

$$q(t) = \left(\begin{smallmatrix} v_1 t \\ c_2 \sin t \end{smallmatrix}\right) \qquad (t \in \mathbb{R}).$$

For $T \in \mathbb{R} \setminus \pi\mathbb{Z}$, the only way to get $q(T) = \binom{c}{0}$ is $v_1 = \frac{c}{T}$ and $c_2 = 0$. In contrast, if $T \in \pi\mathbb{Z}\setminus\{0\}$, then we have the one- parameter family of solutions $q(t) = \left(\frac{c}{T}t, c_2 \sin t\right)^\top$.

(b) The variation of the action functional in direction δq is

$$X(\delta q) = \frac{1}{2}\int_0^T \left[\left\|\begin{pmatrix}\frac{c}{T}\\0\end{pmatrix} + \delta\dot{q}(t)\right\|^2 - \left\|\begin{pmatrix}\frac{c}{T}\\0\end{pmatrix}\right\|^2 - \delta q_2^2(t)\right] \mathrm{d}t$$
$$= \frac{1}{2}\int_0^T [\|\delta\dot{q}(t)\|^2 - \delta q_2^2(t)]\,\mathrm{d}t + \int_0^T \left\langle \begin{pmatrix}\frac{c}{T}\\0\end{pmatrix}, \delta\dot{q}(t)\right\rangle \mathrm{d}t\,.$$

The second integral equals $\left\langle \binom{\frac{c}{T}}{0}, \delta q(T) - \delta q(0)\right\rangle = 0$.

(c) For the given variation, one has $\delta\dot{q}_2(t) = \frac{\pi}{T}\sum_{n=1}^{\infty} n c_n \cos\left(\frac{\pi n t}{T}\right)$, therefore $X(\delta q)$ equals

$$\sum_{n=1}^{\infty} c_n^2 \int_0^T \left[\frac{\pi^2}{T^2}n^2\cos^2\left(\frac{\pi n t}{T}\right) - \sin^2\left(\frac{\pi n t}{T}\right)\right] \mathrm{d}t = \sum_{n=1}^{\infty} \frac{c_n^2}{2T}(\pi^2 n^2 - T^2)\,.$$

(d) For $T \in \big(l\pi, (l+1)\pi\big)$ and $l \in \mathbb{N}_0$, one has $\pi^2 n^2 < T^2$ if and only if $n \le l$. On the subspace determined by $c_k = 0$ $(k > l)$, the quadratic form X is therefore negative definite, but there is no higher dimensional subspace of such variations. □

Exercise 8.21 on page 171 (Tautochrone Problem):
In analogy to (8.3.6), the time that elapses between the start at point (x_0, y_0) and the arrival at the lowest point $(r, -2r)$ is given by

$$\tilde{T}(y_0) := \int_{x_0(y_0)}^{r} \sqrt{\frac{1 + Y'(x)^2}{2g(y_0 - Y(x))}}\,\mathrm{d}x = \int_{x_0}^{r} \sqrt{\frac{r}{gY(x)(Y(x) - y_0)}}\,\mathrm{d}x\,,$$

where we have used that the start velocity is 0; the differential equation (8.3.7) of the brachystochrone has also been used in the manipulation. Substituting the y-variable yields

$$\tilde{T}(y_0) = \int_{y_0}^{-2r} \sqrt{\frac{r}{gy(y - y_0)}}\frac{\mathrm{d}y}{\sqrt{-1 - \frac{2r}{y}}} = \int_{y_0}^{-2r} \sqrt{\frac{r}{g(-y - 2r)(y - y_0)}}\,\mathrm{d}y\,.$$

Substituting $z := \frac{y_0 - y}{2r + y_0}$, hence $y = y_0 - z(2r + y_0)$, yields $\tilde{T}(y_0) = \int_0^1 \sqrt{\frac{r}{gz(1-z)}}\,\mathrm{d}z = \pi\sqrt{\frac{r}{g}}$, which is independent of y_0. □

Exercise 8.23 on page 174 (Length Functional and Euler-Lagrange Equation in Polar Coordinates):

(a) The polar coordinates $(r, \varphi) \in (0, \infty) \times (-\pi, \pi)$ of the plane with a cut are related to cartesian coordinates by $x_1 = r\cos(\varphi)$, $x_2 = r\sin(\varphi)$. Therefore,

$$\dot{x}_1 = \dot{r}\cos\varphi - \dot{\varphi} r\sin\varphi\,,\ \dot{x}_2 = \dot{r}\sin\varphi + \dot{\varphi} r\cos\varphi\,,\quad \text{and}\quad \|\dot{x}\|^2 = \dot{r}^2 + r^2\dot{\varphi}^2.$$

Therefore, the Lagrange function $L(x, \dot{x}) = \frac{1}{2}\|\dot{x}\|^2$ is of the form

$$\tfrac{1}{2}\Big(g_{r,r}(r,\varphi)\dot{r}^2 \;+\; 2g_{r,\varphi}(r,\varphi)\dot{r}\dot{\varphi} \;+\; g_{\varphi,\varphi}(r,\varphi)\dot{\varphi}^2\Big)\,,$$

with $g_{r,r} = 1$, $g_{r,\varphi} = 0$, and $g_{\varphi,\varphi}(r,\varphi) = r^2$. Using that $g^{r,r} = 1$, $g^{r,\varphi} = 0$, and $g^{\varphi,\varphi}(r,\varphi) = r^{-2}$, formula (8.4.2) for the Christoffel symbols yields

$$\Gamma^r_{\varphi,\varphi}(r,\varphi) = -r \quad,\quad \Gamma^\varphi_{r,\varphi}(r,\varphi) = \frac{1}{r} = \Gamma^\varphi_{\varphi,r}(r,\varphi)\,,$$

whereas $\Gamma^r_{r,\varphi} = \Gamma^r_{\varphi,r} = \Gamma^r_{r,r} = \Gamma^\varphi_{\varphi,\varphi} = \Gamma^\varphi_{r,r} = 0$.

(b) With the above form of the metric tensor in polar coordinates, the length functional is of the form $\int_{t_0}^{t_1} \|\dot{x}(t)\|\,\mathrm{d}t = \int_{t_0}^{t_1} \sqrt{\dot{r}^2 + r^2\dot{\varphi}^2}\,\mathrm{d}t$. By (8.4.3), the Euler-Lagrange equations of the energy functional in polar coordinates are $\ddot{r} + \Gamma^r_{\varphi,\varphi}(r,\varphi)\dot{\varphi}^2 = 0$, $\ddot{\varphi} + 2\Gamma^\varphi_{r,\varphi}(r,\varphi)\dot{r}\dot{\varphi} = 0$, hence

$$\ddot{r} = r\dot{\varphi}^2 \quad \text{and} \quad \ddot{\varphi} = -\frac{2}{r}\dot{r}\dot{\varphi}\,. \qquad \square$$

Exercise 8.26 on page 176 (Geodesics on Surfaces of Revolution):
The condition $R(x_3) > 0$ that the profile is positive is important because otherwise, M is not a manifold.

(a) In the formula for the Christoffel symbols,

$$\Gamma^k_{i,j}(x) = \sum_l \tfrac{1}{2} g^{k,l}(x)\Big(\frac{\partial g_{l,j}}{\partial x_i}(x) + \frac{\partial g_{i,l}}{\partial x_j}(x) - \frac{\partial g_{i,j}}{\partial x_l}(x)\Big) \quad \big(i, j, k \in \{r, \varphi\}\big),$$

the metric tensor is $g(x) =$

$$\begin{pmatrix} R'(z)\cos\varphi & R'(z)\sin\varphi & 1 \\ -R(z)\sin\varphi & R(z)\cos\varphi & 0 \end{pmatrix} \begin{pmatrix} R'(z)\cos\varphi & -R(z)\sin\varphi \\ R'(z)\sin\varphi & R(z)\cos\varphi \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} (R'(z))^2+1 & 0 \\ 0 & (R(z))^2 \end{pmatrix},$$

because $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} R(z)\cos\varphi \\ R(z)\sin\varphi \\ z \end{pmatrix}$. Therefore, the Christoffel symbols are

$$\Gamma^z_{z,z} = \tfrac{1}{2} g^{z,z} \frac{\partial g_{z,z}}{\partial z} = \frac{R'(z)R''(z)}{(R'(z))^2+1}, \qquad \Gamma^z_{z,\varphi} = \Gamma^z_{\varphi,z} = -\tfrac{1}{2} g^{z,z} \frac{\partial g_{z,z}}{\partial \varphi} = 0,$$
$$\Gamma^z_{\varphi,\varphi} = -\tfrac{1}{2} g^{z,z} \frac{\partial g_{\varphi,\varphi}}{\partial z} = -\frac{R(z)R'(z)}{(R'(z))^2+1}, \quad \Gamma^\varphi_{z,z} = -\tfrac{1}{2} g^{\varphi,\varphi} \frac{\partial g_{z,z}}{\partial \varphi} = 0,$$
$$\Gamma^\varphi_{z,\varphi} = \Gamma^\varphi_{\varphi,z} = \tfrac{1}{2} g^{\varphi,\varphi} \frac{\partial g_{\varphi,\varphi}}{\partial z} = \frac{R'(z)}{R(z)}, \qquad \Gamma^\varphi_{\varphi,\varphi} = \tfrac{1}{2} g^{\varphi,\varphi} \frac{\partial g_{\varphi,\varphi}}{\partial \varphi} = 0.$$

The general equation for geodesics is $\ddot{x}_k + \sum_{i,j} \Gamma^k_{ij} \dot{x}_i \dot{x}_j = 0$, which in this context translates to the claim.

(b) A meridian $\gamma\colon I \to M$ is of the form

$$\gamma(t) = \begin{pmatrix} R(z(t)) \cos\varphi \\ R(z(t)) \sin\varphi \\ z(t) \end{pmatrix} \text{, hence } \quad \dot{\gamma}(t) = \begin{pmatrix} R'(z(t))\dot{z}(t) \cos\varphi \\ R'(z(t))\dot{z}(t) \sin\varphi \\ \dot{z}(t) \end{pmatrix}.$$

To parametrize by arclength means to require $\|\dot{\gamma}(t)\| = 1$ for all t. This means

$$1 = \|\dot{\gamma}(t)\|^2 = \big(\dot{z}(t)\big)^2 \big((R'(z(t)))^2 + 1\big) \text{, hence } \big(\dot{z}(t)\big)^2 = \big((R'(z(t)))^2 + 1\big)^{-1}. \tag{H.8.1}$$

We observe that $\dot{z}(t) \neq 0$. Taking the derivative yields

$$\dot{z}(t)\ddot{z}(t) = -\big((R'(z(t)))^2 + 1\big)^{-2} R'\big(z(t)\big)\, R''\big(z(t)\big)\, \dot{z}(t).$$

Together with (8.29) and $\dot{\varphi} = 0$, one obtains the geodesic equation for z. The geodesic equation for φ is trivially satisfied.

(c) Parallels of latitude $\gamma\colon \mathbb{R} \to M$ are of the form

$$\gamma(t) = \begin{pmatrix} R(z) \cos\varphi(t) \\ R(z) \sin\varphi(t) \\ z \end{pmatrix} \text{, hence } \quad \dot{\gamma}(t) = \begin{pmatrix} -R(z)\dot{\varphi}(t) \sin\varphi(t) \\ R(z)\dot{\varphi}(t) \cos\varphi \\ 0 \end{pmatrix}.$$

By the geodesic equation for z, and $\dot{z} = \ddot{z} = 0$, it follows that $R'(z) = 0$. By the geodesic equation for φ, the curve γ is parametrized by constant velocity. □

Exercise 8.35 on page 183 (Refraction):

1. We have $\nabla_v L(q, v) = n(q)^2 v$ and $\nabla_q L(q, v) = n(q)\|v\|^2 \nabla n(q)$, hence

$$\frac{\mathrm{d}}{\mathrm{d}t} \nabla_v L(q, \dot{q}) = 2n(q)\, \langle \nabla n(q), \dot{q} \rangle\, \dot{q} + n(q)^2 \ddot{q},$$

and (8.6.1) follows after division by $n > 0$. So L is constant along solutions:

$$\frac{\mathrm{d}}{\mathrm{d}t} L\big(q(t), v(t)\big) = \langle \nabla_v L(q, v), \dot{v} \rangle + \big\langle \nabla_q L(q, v), \dot{q} \big\rangle = 0.$$

Assuming now that n depends only on y, (8.6.2) follows by the substitution $\tilde{y}(x(t)) = y(t)$ of the independent variable x from (8.6.1), as written in the form

$$n(y)\ddot{x} = -2n'(y)\,\dot{x}\dot{y} \quad , \quad n(y)\ddot{y} = n'(y)\,(\dot{x}^2 - \dot{y}^2)\,,$$

because $\ddot{y}(t) = \tilde{y}'\big(x(t)\big)\ddot{x}(t) + \dot{x}(t)^2\tilde{y}''\big(x(t)\big)$.

2. Thus evaluating (8.6.1) for a refractive index n that depends only on the y coordinate yields the differential equation

$$2n'(y)\dot{y}\dot{x} + n(y)\ddot{x} = 0 \quad \text{, or} \quad \frac{\mathrm{d}}{\mathrm{d}t}\big(n^2(y)\dot{x}\big) = 0\,. \tag{H.8.2}$$

Therefore, $\dot{x} = c/n^2(y)$ with some constant c. Integration in (H.8.2) was possible due to the translation invariance of the Lagrangian in x direction.
Substituting $\dot{x} = c/n^2(y)$ into the constant Lagrangian $L = \ell \in \mathbb{R}^+$ yields the relation $1 + (y'(x))^2 = \frac{2\ell}{c^2}n^2(y)$, in other words,

$$y'(x) = \pm\sqrt{\frac{2\ell}{c^2}n^2(y) - 1}\,.$$

This differential equation can be solved by separation of variables. For n of the form (8.6.3), that is $n(y) = n_0 + ky$, one obtains (8.6.4), substituting $Z := \frac{\sqrt{2\ell}}{c}(n_0 + kY)$: $\qquad \pm(x - x_0) =$

$$= \int_{y(x_0)}^{y(x)} \frac{\mathrm{d}Y}{\sqrt{\frac{2\ell}{c^2}(n_0+kY)^2-1}} = \frac{c}{k\sqrt{2\ell}}\int_{z(x_0)}^{z(x)} \frac{\mathrm{d}Z}{\sqrt{Z^2-1}} = \frac{c}{k\sqrt{2\ell}}\,\mathrm{arcosh}(Z)\Big|_{z(x_0)}^{z(x)}\,.$$

3. With the given coordinates for the points a_i, the travel time is

$$T(a_0) = \frac{\|a_1 - a_0\|}{c_1} + \frac{\|a_2 - a_0\|}{c_2} = \frac{\sqrt{x_0^2 + y_1^2}}{c_1} + \frac{\sqrt{(x_2 - x_0)^2 + y_2^2}}{c_2}\,.$$

Denoting this function, which only depends on x_0, as $t(x_0)$, we obtain $t'(x_0) = \frac{\sin\alpha_1}{c_1} - \frac{\sin\alpha_2}{c_2}$, since $\sin\alpha_1 = \frac{x_0}{\sqrt{x_0^2+y_1^2}}$ and $\sin\alpha_2 = \frac{x_2-x_0}{\sqrt{(x_2-x_0)^2+y_2^2}}$. So for a minimal t, one must have $t'(x_0) = 0$; hence Snell's law of refraction. □

Exercise 8.41 on page 185 (Reflection of Light in a Cup):

(a) The rays that are parallel to the 1-axis and hit the circle at $A(\varphi) := \binom{\sin\varphi}{\cos\varphi}$ with $\varphi \in [0, \pi]$ will be reflected at the normal $A(\varphi)$. So the reflected rays have the direction $A\left(\frac{\pi}{2} + 2\varphi\right) = C(2\varphi)$ with $C(\varphi) := \binom{\cos\varphi}{-\sin\varphi}$. On the other hand, using addition theorems of trigonometry, it follows that the reflected ray intersects S^1 the second time in the point $A(3\varphi) = A(\varphi) + 2\sin(\varphi)\,C(2\varphi)$.

(b) From $\mathbb{J}A(\varphi) = -C(\varphi)$ and $\frac{\mathrm{d}}{\mathrm{d}\varphi}A(\varphi) = C(\varphi)$, it follows that

$$\left\langle\frac{\mathrm{d}}{\mathrm{d}\varphi}B_t(\varphi), \mathbb{J}\big(A(3\varphi) - A(\varphi)\big)\right\rangle = \langle tC(\varphi) + 3(1-t)C(3\varphi), C(\varphi) - C(3\varphi)\rangle.$$

For $t = \frac{3}{4}$, this equals $\frac{3}{4}\,\langle C(\varphi) + C(3\varphi), C(\varphi) - C(3\varphi)\rangle = 0$.

(c) The caustic is therefore the curve $\varphi \mapsto B_{3/4}(\varphi)$. It is exactly when $\varphi = \frac{\pi}{2}$ that $\frac{\mathrm{d}}{\mathrm{d}\varphi} B_{3/4}(\varphi) = \frac{3}{4}\big(C(\varphi) + C(3\varphi)\big) = 0$. This corresponds to the point

$$\frac{1}{4}\left(3A\left(\frac{\pi}{2}\right) + A\left(\frac{3\pi}{2}\right)\right) = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}.$$

This is the focus, where the parallel rays will meet. □

Exercise 8.42 on page 187 (Linear Optics):

- First, indeed $N := \begin{pmatrix} 1\!\!1 & \Delta n\, A \\ 0 & 1\!\!1 \end{pmatrix} \in \mathrm{Sp}(4, \mathbb{R})$, because A is symmetric and the condition from Exercise 6.26 (b) applies.
- At the point $\big(O(q), q\big) \in \mathbb{R}^3$ of the interface, the normal points in direction $\binom{1}{-Aq} + \mathcal{O}(\|q\|^3)$. Let the ray in the first medium have direction $v_1 := \binom{1}{p_1/n_1} / \|\binom{1}{p_1/n_1}\|$, then after the refraction in the interface, the direction $v_2 := \binom{1}{p_2/n_2} / \|\binom{1}{p_2/n_2}\|$. By Snell's law, with the unit normal vector $e(q)$ at point $\big(O(q), q\big)$, it follows that the tangential components given by $w_i := v_i - \langle v_i, e(q)\rangle\, e(q)$ are related by $n_1 w_1 = n_2 w_2$. Here we have $e(q) = \binom{1}{-Aq} + \mathcal{O}(\|q\|^2)$, because $\|\binom{1}{-Aq}\| = 1 + \mathcal{O}(\|q\|^2)$. This implies $w_i = v_i - \binom{1}{-Aq} + \mathcal{O}(\|x\|^2) = \binom{0}{p_i/n_i + Aq} + \mathcal{O}(\|x\|^2)$, where $x = (p, q) \in \mathbb{R}^2 \times \mathbb{R}^2$ is the applicable point in phase space. Thus by Snell's law,
$$p_2 = p_1 + \Delta n\, Aq + \mathcal{O}(\|x\|^2) \quad \text{, with} \quad \Delta n := n_1 - n_2\,.$$ □

Exercise 8.45 on page 188 (Optical Devices):

1. The eyepiece, which is the lens of the microscope that is adjacent to the eye, is a thin lens with focal distance f_{eye}; and for the objective, which is the lens facing the object, we denote the corresponding quantity as f_{obj}. In order to have a relaxed view at the object, the rays that leave the eyepiece after coming from a point in the object are to be a family of parallels. We can achieve this by adjusting the distance d between the lenses. With distance d_{obj} between object and objective, one has the following matrix M_d of the microscope:

$$\begin{pmatrix} 1 & -1/f_{\mathrm{eye}} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ d & 0 \end{pmatrix}\begin{pmatrix} 1 & -1/f_{\mathrm{obj}} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ d_{\mathrm{obj}} & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1 + d(d_{\mathrm{obj}} - f_{\mathrm{obj}}) - d_{\mathrm{obj}}(f_{\mathrm{eye}} + f_{\mathrm{obj}})}{f_{\mathrm{eye}} f_{\mathrm{obj}}} & \frac{d - f_{\mathrm{eye}} - f_{\mathrm{obj}}}{f_{\mathrm{eye}} f_{\mathrm{obj}}} \\ d + d_{\mathrm{obj}} - \frac{d d_{\mathrm{obj}}}{f_{\mathrm{obj}}} & 1 - \frac{d}{f_{\mathrm{obj}}} \end{pmatrix}.$$

According to our specifications, the light emanating from a point has to be parallelized by M_d, i.e., $M_d \binom{c}{0} = \binom{0}{c'}$. So the top left matrix entry has to vanish, which happens for distance $d = f_{\mathrm{eye}} + b_{\mathrm{obj}}$ with image distance b_{obj} of the objective. Choosing this parameter in M_d as $d = f_{\mathrm{eye}} + \frac{d_{\mathrm{obj}} f_{\mathrm{obj}}}{d_{\mathrm{obj}} - f_{\mathrm{obj}}}$, it follows that

$$M_d = \begin{pmatrix} 0 & \frac{f_{\text{obj}}}{(d_{\text{obj}}-f_{\text{obj}})f_{\text{eye}}} \\ f_{\text{eye}} - \frac{d_{\text{obj}} f_{\text{eye}}}{f_{\text{obj}}} & 1 - \frac{f_{\text{eye}}}{f_{\text{obj}}} + \frac{d_{\text{obj}}}{f_{\text{obj}} - f_{\text{eye}}} \end{pmatrix}.$$

Multiplying a change of $\binom{0}{\Delta q}$ in a point of the object by M_d results in a change of $m_{22}\Delta q$ to the angle of the outgoing ray with respect to the optical axis. If the object were viewed with bare eye at a distance D (for instance $D = 25$ cm), this change in angle would be $\frac{\Delta q}{D}$. Comparing the two yields the magnification of the microscope to be a factor

$$\frac{d - f_{\text{eye}} - f_{\text{obj}}}{f_{\text{eye}} \cdot f_{\text{obj}}} \cdot D\,.$$

For instance, for focal lengths $f_{\text{obj}} = f_{\text{eye}} = 25$ mm and lense distance $d = 30$ cm, one obtains the object distance $d_{\text{obj}} = 27.5$ mm and magnification by a factor 100.

2. Optometrists prefer to calculate in diopters, which are reciprocals of the focal lengths (with the unit 1 dpt $= 1/\text{m}$). As can be seen in (8.7.5), these quantities add when lenses are combined with distance $d = 0$, because $L_0 = \begin{pmatrix} 1\!\!1 & \left(\frac{1}{f_1} + \frac{1}{f_2}\right) 1\!\!1 \\ 0 & 1\!\!1 \end{pmatrix}$.
For positive distances d, one can read off the *Gullstrand formula* for the total refractivity D_{tot}:

$$D_{\text{tot}} = D_1 + D_2 - d D_1 D_2 \quad , \text{for} \quad D_i := 1/f_i\,.$$

If D_1 is the refractivity of the eye (i.e., of the combination of cornea and lens), and D_{tot} the reciprocal of the distance to the macula, then the eyeglasses needs to have $D_2 = \frac{D_{\text{tot}} - D_1}{1 - d D_1}$ diopters to correct for far view.
This formula also applies separately to the two refractivities of the astigmatic eye, because the main axes of refraction are orthogonal to each other (see also Example 8.24).
Typical values are $D_{\text{tot}} = 60$ dpt, and the distance between eyeglasses and cornea is $d = 14$ mm. For contacts, one has $d = 0$, and the necessary refractivity changes accordingly. □

H.9 Chapter 9, Ergodic Theory

Exercise 9.6 on page 194 (Invariant Measure):
On its intervals of continuity, the Gauss map h is equal to the maps

$$h_n : \left(\tfrac{1}{n+1}, \tfrac{1}{n}\right] \to [0, 1) \quad , \quad x \mapsto \tfrac{1}{x} - n \qquad (n \in \mathbb{N}).$$

These maps have as inverses the maps $f_n : [0, 1) \to \left(\frac{1}{n+1}, \frac{1}{n}\right]$, $y \mapsto \frac{1}{y+n}$. The image (i.e., push-forward) of the measure whose density is $x \mapsto \frac{1}{1+x}$ has therefore the density

$$\sum_{n=1}^{\infty} \frac{|f_n'(y)|^{-1}}{1+f_n(y)} = \sum_{n=1}^{\infty} \frac{1}{(y+n)(y+n+1)} = \sum_{n=1}^{\infty} \left(\frac{1}{y+n} - \frac{1}{y+n+1}\right) = \frac{1}{y+1}$$

at location $y \in [0, 1)$.

Therefore, the measure is invariant under the Gauss map. μ is a probability measure since $\int_0^1 \frac{1}{1+x}\,\mathrm{d}x = \log 2$. □

Exercise 9.10 on page 195 (Phase Space Volume):
For energy $E < 0$, the volume of the domain in phase space below E equals

$$V(E) = \lambda^{2n}\big(H^{-1}\big((-\infty, E]\big)\big) = \int_{B^n_{|E|^{-1/a}}} \int_{B^n_{\sqrt{2(E+\|q\|^{-a})}}} \mathrm{d}p\,\mathrm{d}q$$
$$= v^{(n)} s^{(n-1)} \int_0^{|E|^{-1/q}} r^{n-1}\big(2(E+r^{-a})\big)^{n/2}\,\mathrm{d}r\,,$$

where $v^{(n)} = \lambda^n(B_1^n)$ denotes the Lebesgue measure of n-dimensional unit ball, and $s^{(n-1)}$ denotes the measure of S^{n-1}. At radius $r = 0$, the term under the integral is asymptotic to $2^{n/2}\,r^{n(1-\frac{a}{2})-1}$; therefore, this term is integrable exactly when $a \in (0, 2)$. □

Exercise 9.23 on page 203 (Decay of Correlation): This exercise is based on [BrSi]. We use the orthonormal basis e_k of characters from (9.3.5).

(a) We need to control the expression $\langle f, U^n g\rangle - \langle f, 1\!\!1\rangle\langle 1\!\!1, g\rangle$

$$= \Big\langle \sum_{k\in\mathbb{Z}^2} f_k e_k(x), \sum_{\ell\in\mathbb{Z}^2} g_\ell e_\ell(T^n x)\Big\rangle - \Big\langle \sum_{k\in\mathbb{Z}^2} f_k e_k(x), 1\!\!1\Big\rangle\Big\langle 1\!\!1, \sum_{\ell\in\mathbb{Z}^2} g_\ell e_\ell(x)\Big\rangle$$
$$= \sum_{k,\ell\in\mathbb{Z}^2} f_k \overline{g_\ell}\Big\langle e_k(x), e_{(\hat{T}^\top)^n \ell}(x)\Big\rangle - f_0\overline{g_0} = \sum_{k\in\mathbb{Z}^2\setminus\{0\}} f_k \overline{g_{\tilde{T}^{-n}k}}\,.$$

(b) We view this sum as a scalar product and apply Hölder's inequality, and the last inequality from the hint; the latter is seen by inserting x: for $x \ge y \ge 1$, one has $xy \ge x \ge \frac{x+y}{2}$.

$$\left|\sum_{k\in\mathbb{Z}^2\setminus\{0\}} f_k \overline{g_{\tilde{T}^n k}}\right| = \left|\sum_{k\in\mathbb{Z}^2\setminus\{0\}} (\|k\|_2 f_k)(\|\tilde{T}^n k\|_2 \overline{g_{\tilde{T}^n k}})(\|k\|_2 \|\tilde{T}^n k\|_2)^{-1}\right|$$
$$\le \tfrac{1}{2} L \Big(\sum_{k\in\mathbb{Z}^2\setminus\{0\}} (\|k\|_2 \|\tilde{T}^n k\|_2)^{-4}\Big)^{\frac{1}{4}}$$
$$\le L \Big(\sum_{k\in\mathbb{Z}^2\setminus\{0\}} (\|k\|_2 + \|\tilde{T}^n k\|_2)^{-4}\Big)^{\frac{1}{4}} \qquad \text{(H.9.1)}$$

where $L := 2\big(\sum_{k\in\mathbb{Z}^2\setminus\{0\}} \|k\|_2^2 |f_k|^2\big)^{\frac{1}{2}} \big(\sum_{h\in\mathbb{Z}^2\setminus\{0\}} \|h\|_2^4 |g_h|^4\big)^{\frac{1}{4}}$. We have $L < \infty$ since the ℓ^p spaces are monotonic with respect to p, in particular $\ell^2(\mathbb{Z}^2) \subset \ell^4(\mathbb{Z}^2)$. The claim follows.

(c) • First consider even $n = 2m$. We have to estimate the terms $\|k\|_2 + \|\tilde{T}^n k\|_2$ in (H.9.1) from below, and we use $h := \tilde{T}^m k$ to this end:

$$\begin{aligned}\|\tilde{T}^m h\|_2 + \|\tilde{T}^{-m} h\|_2 &\ge C^{-1}(\|\tilde{T}^m h\|_E + \|\tilde{T}^{-m} h\|_E)\\ &= C^{-1}(\lambda^{-m}\|h_s\|_2 + \lambda^m \|h_u\|_2 + \lambda^m \|h_s\|_2 + \lambda^{-m}\|h_u\|_2)\\ &\ge C^{-1}\lambda^m \|h\|_E \ge C^{-2}\lambda^m \|h\|_2 .\end{aligned}$$

As the mapping $k \mapsto h = \tilde{T}^m k$ is a permutation of $\mathbb{Z}^2 \setminus \{0\}$, we obtain

$$|\langle f, U^n g\rangle - \langle f, 1\rangle\langle 1, g\rangle| \le C^2 L \lambda^{-m} \Big(\sum_{h\in\mathbb{Z}^2\setminus\{0\}} \|h\|_2^{-4}\Big)^{\frac{1}{4}}.$$

• For odd $n = 2m' + 1$, we obtain analogously

$$\begin{aligned}\|\tilde{T}^{m'+1} h\|_2 + \|\tilde{T}^{-m'} h\|_2 &\ge C^{-1}(\|\tilde{T}^{m'+1} h\|_E + \|\tilde{T}^{-m'} h\|_E)\\ &= C^{-1}(\lambda^{-m'-1}\|h_s\|_2 + \lambda^{m'+1}\|h_u\|_2 + \lambda^{m'}\|h_s\|_2 + \lambda^{-m'}\|h_u\|_2)\\ &\ge C^{-1}\lambda^{m'}\|h\|_E \ge C^{-2}\lambda^{m'}\|h\|_2 ,\end{aligned}$$

hence $|\langle f, U^n g\rangle - \langle f, 1\rangle\langle 1, g\rangle| \le C^2 L \lambda^{-m'} \Big(\sum_{h\in\mathbb{Z}^2\setminus\{0\}} \|h\|_2^{-4}\Big)^{\frac{1}{4}}$.

• From $m = \frac{n}{2} = \lfloor \frac{n}{2} \rfloor$ and $m' = \frac{n-1}{2} = \lfloor \frac{n}{2} \rfloor$, the claim follows. □

Exercise 9.25 on page 205 (Product Measure on the Shift Space):

• The cylinder sets form a family $\mathcal{S}$ of sets that generates the σ-algebra $\mathcal{M}$. First, if $A, B \in \mathcal{S}$, then there exists $T \in \mathbb{N}$ with

$$\mu_p(\Phi_t(A) \cap B) = \mu_p(A)\,\mu_p(B) \qquad (|t| \ge T)$$

for the product measures μ_p. For the cylinder sets $A = [\tau_1, \ldots, \tau_j]_k^{k+j-1}$ and $B = [\kappa_1, \ldots, \kappa_i]_\ell^{\ell+i-1}$, this occurs for $T := \max(\ell + i - k,\ k + j - \ell)$.

- Now if $A, B \in \mathcal{M}$, then for all $\varepsilon > 0$, there exist elements $\tilde{A}, \tilde{B}$ of the algebra $\mathcal{A}(\mathcal{S})$ generated by the cylinder sets such that the symmetric differences satisfy $\mu_p(\tilde{A} \triangle A) < \varepsilon$ and $\mu_p(\tilde{B} \triangle B) < \varepsilon$ (see for instance ELSTRODT [El]). For this pair, there also exists a minimal time $\tilde{T}$ such that

$$\mu_p\big(\Phi_t(\tilde{A}) \cap \tilde{B}\big) = \mu_p(\tilde{A})\,\mu_p(\tilde{B}) \qquad \big(|t| \geq \tilde{T}\big).$$

On the other hand, since

$$\big(\Phi_t(A) \cap B\big) \triangle \big(\Phi_t(\tilde{A}) \cap \tilde{B}\big) \subseteq \big(\Phi_t(A) \triangle \Phi_t(\tilde{A})\big) \cup \big(B \triangle \tilde{B}\big)$$

for all times t, the measure of the left hand side is less than 2ε. Now if $|t| \geq \tilde{T}$, then it follows that

$$\begin{aligned}
&\big|\mu_p\big(\Phi_t(A) \cap B\big) - \mu_p(A)\,\mu_p(B)\big| \leq \\
&\quad \big|\mu_p\big(\Phi_t(A) \cap B\big) - \mu_p\big(\Phi_t(\tilde{A}) \cap \tilde{B}\big)\big| + \big|\mu_p\big(\Phi_t(\tilde{A}) \cap \tilde{B}\big) - \mu_p(\tilde{A})\,\mu_p(\tilde{B})\big| \\
&\quad + |\mu_p(\tilde{A}) - \mu_p(A)|\,\mu_p(\tilde{B}) \;+\; \mu_p(A)\,|\mu_p(\tilde{B}) - \mu_p(B)| \\
&\quad < 2\varepsilon + 0 + \varepsilon + \varepsilon = 4\varepsilon\,.
\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, it follows that $\lim_{|t|\to\infty} \mu_p\big(\Phi_t(A) \cap B\big) = \mu_p(A)$ $\mu_p(B)$. □

Exercise 9.27 on page 207 (Shift Space):

- We begin by finding *a single* point $m = \{m_k\}_{k\in\mathbb{Z}} \in M$ for which the averages have the interval $[-1, 1]$ as the set of cluster points. To this end, let

$$m_k := \begin{cases} 1 & , k \in \{-1, 0, 1\} \\ (-1)^{\lfloor \log_2(\log_2 |k|)\rfloor} & , k \in \mathbb{Z} \setminus \{-1, 0, 1\}\,. \end{cases}$$

Now we choose the subsequence $n_a := 2^{(2^a)}$ and note for $a \in \mathbb{N}_0$ that

$$\begin{aligned}
k &\in \{n_{2a}, \ldots, n_{2a+1} - 1\} && \Longrightarrow && m_k = 1 \\
k &\in \{n_{2a+1}, \ldots, n_{2a+2} - 1\} && \Longrightarrow && m_k = -1\,.
\end{aligned}$$

This allows us to estimate the averages for the subsequence:

$$\begin{aligned}
|A_{n_{2a}} f(m) + 1| &= \Big|\tfrac{1}{n_{2a}} \textstyle\sum_{t=0}^{n_{2a}-1} f \circ \Phi^t(m) + 1\Big| \\
&\leq \frac{1}{n_{2a}} \sum_{t=0}^{n_{2a-1}-1} 1 + \Big|\frac{1}{n_{2a}} \sum_{t=n_{2a-1}}^{n_{2a}-1} f \circ \Phi^t(m) + 1\Big| \\
&\leq \tfrac{n_{2a-1}}{n_{2a}} + \Big|\tfrac{n_{2a}-n_{2a-1}}{n_{2a}} - 1\Big| \leq 2\tfrac{n_{2a-1}}{n_{2a}} = 2^{1+2^{a-1}-2^a} = 2^{1-2^{a-1}} \xrightarrow{a\to\infty} 0\,,
\end{aligned}$$

i.e., $\lim_{a\to\infty} A_{n_{2a}} f(m) = -1$. Likewise we obtain $\lim_{a\to\infty} A_{n_{2a+1}} f(m) = 1$. This shows that ± 1 are cluster points of the sequence $\big(A_n f(m)\big)_{n\in\mathbb{N}}$. Since

$$|A_n f(m) - A_{n+1} f(m)| = \Big|\frac{1}{n}\sum_{t=0}^{n-1} f\circ T^t(m) - \frac{1}{n+1}\sum_{t=0}^{n} f\circ T^t(m)\Big|$$

$$\le \Big(\frac{1}{n} - \frac{1}{n+1}\Big)\sum_{t=0}^{n-1} |f\circ T^t(m)|$$

$$+\frac{1}{n+1}\Big|\sum_{t=0}^{n-1} f\circ T^t(m) - \sum_{t=0}^{n} f\circ T^t(m)\Big| \le \frac{2}{n+1},$$

all points between -1 and 1 must be cluster points, too.

- Now we have to extend $\{m\} \subseteq M$ to a dense subset U of M. We define for all $m' \in M$:

$$x^{m',s} \in M \quad \text{with} \quad x_k^{m',s} := \begin{cases} m'_k & , |k| \le s \\ m_k & , |k| > s \end{cases} \qquad (s\in\mathbb{N}, k\in\mathbb{Z}).$$

Thus $\lim_{s\to\infty} x^{m',s} = m'$. The set $U := \big\{x^{m',s} \in M \mid m' \in M,\ s\in\mathbb{N}\big\}$ is therefore dense in M and has the desired property of the cluster points. □

Exercise 9.35 on page 211 (Birkhoff's Ergodic Theorem for Flows):

- The flow $\Phi : \mathbb{R}\times M \to M$ is continuous, hence measurable. The time-t mappings $\Phi_t : M \to M$ $(t\in\mathbb{R})$ preserve the measure μ. So we can apply Fubini's theorem for all $T > 0$ and for the restriction λ_T of the Lebesgue measure to $[0, T]$; $\int_{[0,T]\times M} f\circ\Phi \,\mathrm{d}\lambda_T\,\mathrm{d}\mu = T\int_M f\,\mathrm{d}\mu = \int_M\big(\int_{[0,T]} f\circ\Phi(t,m)\,\mathrm{d}\lambda(t)\big)\mathrm{d}\mu$. In particular, the inner integral exists for μ-almost all $m\in M$.
- For μ-almost all initial conditions $m \in M$, we can estimate the integral over time $\int_0^T f\circ\Phi(t,m)\,\mathrm{d}t$ by the one whose upper limit is an integer, because, letting $G := \int_0^1 |f\circ\Phi_t|\,\mathrm{d}t \in L^1(M,\mu)$, one has

$$\Big|\int_0^T f\circ\Phi(t,m)\,\mathrm{d}t - \int_0^{\lfloor T\rfloor} f\circ\Phi(t,m)\,\mathrm{d}t\Big| \le G\circ\Phi\big(\lfloor T\rfloor, m\big)\,. \qquad \text{(H.9.2)}$$

Since by Birkhoff's ergodic theorem (Theorem 9.32), $\overline{G}(m)$ exists for μ-almost all $m\in M$, the proof of Lemma 9.28 also yields

$$\lim_{n\to\infty} \tfrac{1}{n} G\circ\Phi_n(m) = 0 \qquad (\mu\text{-almost everywhere}).$$

Therefore, from (H.9.2) and the ergodic theorem, it follows μ-almost everywhere:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f\circ\Phi(t,m)\,\mathrm{d}t \;=\; \lim_{T\to\infty}\frac{1}{\lfloor T\rfloor}\int_0^{\lfloor T\rfloor} f\circ\Phi(t,m)\,\mathrm{d}t \;=\; \overline{F}(m)$$

with $F := \int_0^1 f \circ \Phi_t \, dt \in L^1(M, \mu)$. Therefore, $\overline{f} = \overline{F}$ holds μ-almost everywhere. The remaining claims about $\overline{f}$ follow by the ergodic theorem and Remark 9.33.1 from the analogous claims for $\overline{F}$. □

Exercise 9.36 on page 212 (Normal Real Numbers):

- The mapping

$$T : M \to M \quad , \quad x \mapsto 2x \pmod 1 \quad \text{on } M := [0, 1)$$

  is ergodic, and even mixing with respect to the restriction to $M \subseteq \mathbb{R}$ of the T-invariant Lebesgue measure λ. This is because $L^2(M, \lambda)$ is spanned by the functions $e_k : M \to S^1$, $e_k(x) = \exp(2\pi \imath k x)$ $(k \in \mathbb{Z})$, and $e_k\big(T(x)\big) = e_{2k}(x)$.
- The set $N \subseteq M$ of numbers whose dyadic expansion is not unique (namely those whose period is either $\overline{0}$ or $\overline{1}$) is countable, hence its Lebesgue measure $\lambda(N) = 0$. Moreover $T(N) = N$.
- Since T acts, in the dyadic representation, as a shift by one digit, we consider the function $f := \mathbb{1}_{[0, \frac{1}{2})} \in L^1(M, \lambda)$.

By Birkhoff's ergodic theorem (Theorem 9.32), the time average $\overline{f}(x)$ of f exists for λ-almost all $x \in M$. According to the remark above, it can be interpreted as a frequency for λ-almost all $x \in M$. By the ergodicity of T, it equals, for λ-almost all $x \in M$, the space average $\int_M f \, d\lambda = \frac{1}{2}$. □

H.10 Chapter 10, Symplectic Geometry

Exercise 10.8 on page 220 (Particles in a Magnetic Field):

(a) • The form $\omega_B = \omega_0 + B_1 \, dq_2 \wedge dq_3 + B_2 \, dq_3 \wedge dq_1 + B_3 \, dq_1 \wedge dq_2 \in \Omega^2(P)$ on the phase space $P = \mathbb{R}^3_q \times \mathbb{R}^3_v$ with the symplectic form $\omega_0 = \sum_{i=1}^3 dq_i \wedge dv_i$ is *closed* because of the hypothesis div $(B) = 0$ and

$$\begin{aligned} d\omega_B &= dB_1 \wedge dq_2 \wedge dq_3 + dB_2 \wedge dq_3 \wedge dq_1 + dB_3 \wedge dq_1 \wedge dq_2 \\ &= \operatorname{div}(B) \, dq_1 \wedge dq_2 \wedge dq_3 \, . \end{aligned}$$

• ω_B is *non-degenerate* because, given a tangent vector $X \neq 0$ at $(q, v) \in P$, there exists some $i \in \{1, 2, 3\}$ for which the $\frac{\partial}{\partial q_i}$ component of X doesn't vanish, or else, the $\frac{\partial}{\partial v_i}$-component of X doesn't.
In the first case, we set $Y := \frac{\partial}{\partial v_i}$, in the second case $Y := \frac{\partial}{\partial q_i}$. In either case, $\omega_B(X, Y) = \omega_0(X, Y) \neq 0$.

(b) With $X = \sum_{i=1}^3 \Big(X_i^{(v)} \frac{\partial}{\partial v_i} + X_i^{(q)} \frac{\partial}{\partial q_i} \Big)$ and $Y = \sum_{i=1}^3 \Big(Y_i^{(v)} \frac{\partial}{\partial v_i} + Y_i^{(q)} \frac{\partial}{\partial q_i} \Big)$, one has $\omega_B(X, Y) = \big\langle X^{(q)}, Y^{(v)} \big\rangle - \big\langle X^{(v)}, Y^{(q)} \big\rangle + \det(B, X^{(q)}, Y^{(q)})$.

(c) $dH(q,v) = \sum_{i=1}^{3} v_i dv_i$, hence $dH(Y)(q,v) = \langle v, Y^{(v)}(q,v)\rangle$. Comparing coefficients in $\omega_B(X_H,\cdot) = dH$ and using $\omega_B(X,Y) = \langle X^{(q)}, Y^{(v)}\rangle + \langle B \times X^{(q)} - X^{(v)}, Y^{(q)}\rangle$, one obtains $X^{(q)}(q,v) = v,\ X^{(v)}(q,v) = B(q) \times v$.

(d) From (c), one obtains the equations of motion under the Lorentz force (see (6.3.14)):

$$\dot{q} = v \quad , \quad \dot{v} = B(q) \times v$$

□

Exercise 10.30 on page 229 (Sphere and Cylinder):
The total derivative of the mapping $F : \mathcal{Z} \to S^2$ at position $x \in \mathcal{Z}$ is

$$\mathrm{D}F_x = \begin{pmatrix} w & 0 & -\frac{x_1 x_3}{w} \\ 0 & w & -\frac{x_2 x_3}{w} \\ 0 & 0 & 1 \end{pmatrix},$$

where the abbreviation $w := \sqrt{1 - x_3^2} > 0$ was used for the square root.

The area element φ on the cylinder at $x \in \mathcal{Z}$ is given by

$$\varphi_x(Y, Z) = \det(\tilde{x}, Y, Z) \qquad (Y,\ Z \in T_x\mathcal{Z}),$$

where $\tilde{x} := \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$ is the radial component of x. This yields

$$\begin{aligned} \omega_{F(x)}\big(\mathrm{D}F_x(Y)\,,\,\mathrm{D}F_x(Z)\big) &= \det\begin{pmatrix} x_1 w & Y_1 w - Y_3\frac{x_1 x_3}{w} & Z_1 w - Z_3\frac{x_1 x_3}{w} \\ x_2 w & Y_2 w - Y_3\frac{x_2 x_3}{w} & Z_2 w - Z_3\frac{x_2 x_3}{w} \\ x_3 & Y_3 & Z_3 \end{pmatrix} \\ &= x_1(Y_2 Z_3 - Y_3 Z_2) + x_2(Y_3 Z_1 - Y_1 Z_3) = \varphi_x(Y, Z)\,. \end{aligned}$$

In calculating the determinant, we have used that for tangent vectors Y, Z to the cylinder, the 3-component $Y_1 Z_2 - Y_2 Z_1$ of the cross product is 0. □

Exercise 10.45 on page 240 (Representation of the Flow by Generating Functions):

(a) The generating function $H : (-\pi/2, \pi/2) \times \mathbb{R}^2 \to \mathbb{R}$ also is continuous with respect to time, because $\lim_{t\to 0} H_t(p,q) = H_0(p,q)$. According to (10.5.1),

$$p_t = \frac{p_0}{\cos(t)} - q_t \tan(t) \quad , \quad q_t = q_0 + q_t\left(1 - \frac{1}{\cos(t)}\right) + p_0 \tan(t)\,,$$

hence $q_t = q_0\cos(t) + p_0\sin(t),\ \ p_t = p_0\cos(t) - q_0\sin(t)$.
This is the solution to the differential equation for the harmonic oscillator with Hamiltonian H_0.

(b) The solution to the Hamiltonian equations for the quadratic Hamiltonian $H_0(x) = \frac{1}{2}\langle x, Ax\rangle$ and initial value $x_0 = (p_0, q_0)$ is given as $x_t = (p_t, q_t) = \exp(\mathbb{J}At)x_0$, hence is linear.
Therefore, there exists a time $T > 0$ such that for all $|t| < T$, the position at t as a function of the initial values has the property that $\mathrm{D}_2 q_t(p_0, q_0)$ is of maximal

rank (i.e., rank n). By the implicit function theorem, we can invert the relation and find a smooth mapping $Q_t : \mathbb{R}^{2n} \to \mathbb{R}^n$, $q_0 = Q_t(p_0, q_t)$.
The generating function H_t satisfies the integrable relation $\mathrm{D}H_t(p_0, q_t) = \mathbb{J}\left(\frac{x_0 - x_t}{t}\right)$, whose right hand side is known. H_t is determined by the convention $H_t(0) = 0$. As $x_t = \exp(\mathbb{J}At)x_0$, it follows that $\lim_{t\to 0} \frac{x_t - x_0}{t} = \mathbb{J}Ax_0$, hence $\lim_{t\to 0} \mathrm{D}H_t(x) = Ax_0$. This in turn can be integrated to $H_0(x) = \frac{1}{2}\langle x, Ax\rangle$.

(c) The anharmonic oscillator has a Hamiltonian

$$H_0 : \mathbb{R}^2 \to \mathbb{R} \quad , \quad H_0(p, q) = \tfrac{1}{2}p^2 + V(q) \quad \text{with} \quad V(q) := \tfrac{1}{2}(q^2 + q^4) \geq 0.$$

By Theorem 11.1, H_0 generates a flow $\Phi \in C^1\left(\mathbb{R} \times \mathbb{R}^2, \mathbb{R}^2\right)$. As $V'(q) = q + 2q^3$ is an odd function of the position q, with $V'(q) > 0$ for $q > 0$, the point $(p_0, q_0) = (0, 0)$ is the only equilibrium.
Since moreover, $\lim_{|q|\to\infty} V(q) = +\infty$, the energy curves $H_0^{-1}(h) \subset \mathbb{R}^2$ are diffeomorphic to the circle S^1 for all $h > 0$. So all orbits are periodic. The period is a function $T : (0, \infty) \to (0, \infty)$ of the energy h. We obtain $\lim_{h\to\infty} T(h) = 0$ because, denoting the maximal elongation as $c(h) := \sqrt{\frac{\sqrt{1+8h}-1}{2}}$ and letting $Q := q/c(h)$, we have the formula for $T(h)$: it equals

$$\int_{-c(h)}^{c(h)} \frac{\mathrm{d}q}{\mathrm{d}q/\mathrm{d}t} = 2\int_0^{c(h)} \frac{\mathrm{d}q}{\sqrt{2h - q^2 - q^4}} = \frac{2}{c(h)} \int_0^1 \frac{\mathrm{d}Q}{\sqrt{\frac{2h}{c(h)^4} - \left(\frac{Q}{c(h)}\right)^2 - Q^4}}.$$

As $h \to \infty$, the limit of the integral is $\int_0^1 \frac{\mathrm{d}Q}{\sqrt{1-Q^4}} > 0$, which implies the claim $\lim_{h\to\infty} T(h) = 0$. But this proves that (in contrast to (b)), for the *an*harmonic oscillator, the implicit equation $q_0 = Q_t(p_0, q_t)$ cannot be solved on the entire phase space for times t in any interval $(-T, T)$. □

H.11 Chapter 11, Motion in a Potential

Exercise 11.2 on page 243 (Going to Infinity in Finite Time):

- If $c \geq 0$, then $V \in C^\infty(\mathbb{R}^d, \mathbb{R})$ is nonnegative. Therefore, by Theorem 11.1, the Hamiltonian differential equation (11.1.2) generates a flow.
- For $c < 0$ and a solution $x = (p, q) : I \to P$, we consider the function $f : I \to [0, \infty)$, $f(t) = 1 + \|q(t)\|^2$. The energy is constant in time, and for initial conditions (p_0, q_0) with $p_0 = \sqrt{2(E - V(q_0))}\frac{q_0}{\|q_0\|} \neq 0$, one has $H(x(t)) = E$. Moreover, $\dot{p} = -2c(1 + \varepsilon)(1 + \|q\|^2)^\varepsilon q$, and therefore $p(t)$ and $q(t)$ lie in span(q_0), and their absolute values increase with time. So if $E \geq 0$, then f satisfies the differential **in**equality

$$f'(t) = 2\langle q(t), p(t)\rangle = 2\sqrt{2(E + |c|(1 + \|q(t)\|^2)^{1+\varepsilon})}\|q(t)\| \geq kf(t)^{1+\varepsilon/2}$$

with $k := \sqrt{|c|}\min(1, \|q_0\|)$, because $\|q(t)\|^2 \geq \frac{1}{2}(1 + \|q(t)\|^2)\min(1, \|q_0\|^2)$. The corresponding differential **equation** $g'(t) = kg(t)^{1+\varepsilon/2}$ has the solution $g(t) = \left(g(0) - k\frac{\varepsilon}{2}t\right)^{-\frac{2}{\varepsilon}}$, so it diverges at time $\frac{2g(0)}{\varepsilon k}$. Letting $g(0) := f(0)$, it follows that for $t \in I,\ t \geq 0$, one has $f(t) \geq g(t)$ as well. This implies the divergence of f in finite time.

A modification of this argument shows that solutions for energy $E < 0$ also diverge. □

Exercise 11.15 on page 253 (Ballistic and Bound Motion):

(a) We shorten the closed curve $c_\ell : S^1 \to \mathbb{T},\ c_\ell(t) = t\,\ell$ in the Jacobi metric and obtain a closed geodesic $\widehat{c_\ell} : S^1 \to \mathbb{T}$ as in Theorem 8.33.
According to Theorem 8.31, this geodesic corresponds to a time-periodic solution curve $t \mapsto \widehat{q}(t, \hat{x}_0)$ with initial value $\hat{x}_0 \in \widehat{H}^{-1}(E)$ and period $T > 0$.
The claim of the theorem applies to every initial condition $x_0 \in \Sigma_E$ with projection $\pi(x_0) = \hat{x}_0$.

(b) Let $\tilde{V} \in C^2\big([0, \infty), (-\infty, 0]\big)$ be a real function with $\tilde{V}'(0) = 0$ and $\tilde{V}(r) = 0$ for all $r \geq R > 0$. Then for dimension $d \in \mathbb{N}$ and the lattice $\mathcal{L} := 2R\mathbb{Z}^d$, the $\mathcal{L}$-periodic potential

$$V : \mathbb{R}^d \to \mathbb{R} \quad , \quad V(q) = \sum_{\ell \in \mathcal{L}} \tilde{V}(\|q - \ell\|)$$

is also twice continuously differentiable. Moreover, $V(q) = \tilde{V}(\|q\|)$ if $\|q\| \leq R$. As can be seen by considering the effective potential (see page 328), for $d \geq 2$ and appropriately chosen $\tilde{V}$ and $E > 0$, there exist initial conditions $x_0 = (p_0, q_0) \in \Sigma_E$ such that $\|q(t, x_0)\| \leq R$ for all $t \in \mathbb{R}$. The set of these initial conditions has even positive Liouville measure. □

Exercise 11.19 on page 255 (Total Integrability):

Write the separable potential as $V(q) = \sum_{i=1}^d V_i(q_i)$ with T_i-periodic functions V_i. Then the regular lattice $\mathcal{L} := \{(k_1T_1, \ldots, k_dT_d)^\top \mid k_i \in \mathbb{Z}\} \subset \mathbb{R}^d$ is the period lattice of V. For initial value $x_0 = (p_0, q_0) = (p_{1,0}, \ldots, p_{d,0}, q_{1,0}, \ldots, q_{d,0})$, the solution has then the form $(p(t), q(t))$, where $(p_i(t), q_i(t))$ solves the Hamiltonian equations for $H_i : \mathbb{R}^2 \to \mathbb{R},\ H_i(p_i, q_i) = \frac{1}{2}p_i^2 + V_i(q_i)$.

For $d \geq 2$, total energy $E > V_{\max} = \sum_{i=1}^d V_{i,\max}$, and $j \in \{1, \ldots, d\}$, we can choose initial conditions $\big(p^{(0)}, q^{(0)}\big) \in \mathbb{R}^d_p \times \mathbb{R}^d_q$ with $H_j\big(p_j^{(0)}, q_j^{(0)}\big) = V_{j,\max}$ and $H_i\big(p_i^{(0)}, q_i^{(0)}\big) > V_{i,\max}$ for all $i \in \{1, \ldots, d\} \setminus \{j\}$. If the term V_j in the sum is not constant, we may assume $p_j^{(0)} > 0$. In this case, even though $t \mapsto q_j(t)$ is strictly increasing, it is bounded. This is incompatible with a conditionally periodic, hence non-wandering[33] motion on a torus.

Remark: While the motion in such a separable potential is not completely integrable, those points in phase space that do not lie on invariant tori form a null set. Values

[33]**Definition:** A point in a topological dynamical system $\Phi : G \times M \to M$ is called *wandering* if it has a neighborhood $U \subset M$ such that $U \cap \Phi_t(U) = \emptyset$ for all $t \geq T$.

of the energies H_j that are strictly smaller than $V_{j,\max}$ correspond to invariant phase space tori that do not project diffeomorphically to the configuration torus $\mathbb{R}^d/\mathcal{L}$. □

Exercise 11.21 on page 256 (Constants of Motion):

- The total momentum p_N satisfies

$$\dot{p}_N = \sum_{k=1}^{n} \dot{p}_k = -\sum_{k=1}^{n} \nabla_{q_k} V(q) = \sum_{k=1}^{n}\sum_{\ell\neq k} \frac{m_k m_\ell}{\|q_k - q_\ell\|^3}(q_\ell - q_k) = 0\,.$$

- $q_N(p,q;t) = \frac{1}{m_N}\sum_{k=1}^n (m_k\, q_k - p_k\, t)$, the 'center of mass at time 0' on extended phase space, depends in fact explicitly on time t, but

$$\dot{q}_N = \frac{1}{m_N}\sum_{k=1}^{N}(m_k\dot{q}_k - \dot{p}_k t - p_k) = \frac{1}{m_N}\sum_{k=1}^{N}\left(m_k\frac{p_k}{m_k} + \nabla_{q_k} V(q) - p_k\right)$$

$= -\frac{\dot{p}_N}{m_N} = 0.$

- The components of the total angular momentum are constant as well, because

$$\begin{aligned}\dot{L}_{i,j} &= \sum_{k=1}^{n}\big(\dot{q}_{k,i}\, p_{k,j} + q_{k,i}\,\dot{p}_{k,j} - \dot{q}_{k,j}\, p_{k,i} - q_{k,j}\,\dot{p}_{k,i}\big)\\ &= \sum_{k=1}^{n}\left(\frac{1}{m_k} p_{k,i}\, p_{k,j} - q_{k,i}\,\partial_{q_{k,j}} V(q) - \frac{1}{m_k}\, p_{k,j}\, p_{k,i} + q_{k,j}\,\partial_{q_{k,i}} V(q)\right)\\ &= \sum_{k=1}^{n}\sum_{\ell\neq k}\frac{m_k m_\ell}{\|q_k - q_\ell\|^3}\big[q_{k,i}(q_{\ell,j} - q_{k,j}) - q_{k,j}(q_{\ell,i} - q_{k,i})\big] = 0\,.\end{aligned}$$

- The choice of constants $p_N = 0,\ q_N = 0$ implies $\sum_{k=1}^n m_k q_k = 0$. Since these single equations are linearly independent, we have thus defined a $2(n-1)d$-dimensional submanifold of the phase space $\widehat{P}$.
- For one particle ($n = 1$), this submanifold is a point, and there are exactly $2d$ algebraically independent constants of motion.
- In the case of $n = 2$ particles, we can pass to relative coordinates $(p_r, q_r) \in T^*\big(\mathbb{R}^d\backslash\{0\}\big)$, and in these, the motion of a particle is like the motion in the field of a central force. Its Hamiltonian is $H_r(p_r, q_r) = \frac{\|p_r\|^2}{2m_r} - \frac{m_1 m_2}{\|q_r\|}$, with $m_r = \frac{m_1 m_2}{m_1+m_2}$ being the reduced mass (see Example 12.39).
  In $d = 2$ degrees of freedom, we have (dropping the index r):

$$
\begin{aligned}
(dH \wedge dL)(p,q) &= \left(\frac{p_1\,dp_1 + p_2\,dp_2}{m} + \frac{m_1 m_2}{\|q\|^3}(q_1\,dq_1 + q_2\,dq_2)\right) \\
&\wedge (q_1\,dp_2 + p_2\,dq_1 - q_2\,dp_1 - p_1\,dq_2) \\
&= \langle q, p\rangle \left(dp_1 \wedge dp_2 - \frac{m_1 m_2}{\|q\|^3} dq_1 \wedge dq_2\right) + \sum_{i,j=1}^{2} f_{i,j}(p,q)dq_i \wedge dp_j \,.
\end{aligned}
$$

So the two constants of motion H and L are algebraically independent.
A similar consideration for $d = 3$ also yields that H and the $\binom{d}{2}$ components $L_{i,j}$ $(1 \le i < j \le d)$ of the angular momentum are algebraically independent. A fortiori, the same applies for $n > 2$.
We have $2d + 1 + \binom{d}{2} = \binom{d+2}{2}$. So for $d = 2$, we have $\binom{d+2}{2} = 6$ independent constants of motion, and for $d = 3$ we have $\binom{d+2}{2} = 10$ of them. □

Exercise 11.22 on page 258 (Laplace-Runge-Lenz Vector):

(a) The time derivative of the Laplace-Runge-Lenz vector is

$$
\frac{\mathrm{d}}{\mathrm{d}t}\hat{A}(x) = Z\left(\frac{\hat{L}(x)}{\|q\|^3}\begin{pmatrix}-q_2\\ q_1\end{pmatrix} - \frac{\|q\|^2 p - \langle p, q\rangle\, q}{\|q\|^3}\right) = 0\,,
$$

with $x(t) = \big(p(t), q(t)\big)$.
The pericenter q of the orbit is distinguished by satisfying $\langle p, q\rangle = 0$ and at the same time having a positive radial acceleration:

$$
\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\|q(t)\|^2 \;=\; \|p(t)\|^2 - \frac{Z}{\|q(t)\|} \;\ge\; 0\,.
$$

Then $\big\langle \hat{A}(x), q\big\rangle \ge 0$, and $\hat{A}$ is parallel to q.

(b) At the pericenter (and therefore due to (a) along the entire orbit),

$$
\frac{\|\hat{A}\|}{Z} = \frac{\big\langle \hat{A}, q\big\rangle}{Z\|q\|} = \frac{\ell^2/Z - \|q\|}{\|q\|} = e\,,
$$

because the eccentricity is $e = \frac{\ell^2}{ZR} - 1$ according to 1.7. □

Exercise 11.25 on page 263 (Regularizable Singular Potentials):

1. Denoting the angular momentum as $\ell > 0$ and substituting $r = \ell^{\frac{2}{2-a}} R$, the angle of deflection as viewed from the origin can, for the a-homogenous potential, be written as

$$
\Delta\varphi(E,\ell) = 2\int_{R_{\min}}^{\infty} \frac{\mathrm{d}R}{R\sqrt{2E\ell^{\frac{2a}{2-a}} + 2ZR^{2-a} - 1}}\,.
$$

Therefore, $\Delta\varphi(E, 0_+) = 2\int_{R_{\min}}^{\infty} \frac{\mathrm{d}R}{R\sqrt{2ZR^{2-a}-1}}$, with $R_{\min} = (2Z)^{-\frac{1}{2-a}}$.

The substitution $R = (2Z)^{-\frac{1}{2-a}} u$ shows that the angle of deflection is independent not only of E, but also of the charge $Z > 0$, because $\Delta\varphi(E, 0_+) = 2\int_{u_{\min}}^{\infty} \frac{\mathrm{d}u}{u\sqrt{u^{2-a}-1}}$ with $u_{\min} = 1$. Therefore, $\Delta\varphi(E, 0_+) = \frac{2\pi}{2-a}$.
$\Delta\varphi(E, \ell)$ is odd in ℓ, therefore $\lim_{\ell \nearrow 0} \Delta\varphi(E, \ell) = -\frac{2\pi}{2-a}$. These two angles are equal modulo 2π if and only if $a = 2(1 - 1/n)$ for some $n \in \mathbb{N}$.

2. According to Remark 11.24.4, we have to parametrize the orbits of the geodesic flow Ψ on the unit tangent bundle $T_1 S^d$.
   (a) For $d = 2$, this can be done by the mapping

$$T_1 S^2 \to S^2 \quad , \quad (x, y) \mapsto x \times y\,.$$

   This mapping is Ψ-invariant, because, letting $\big(x(t), y(t)\big) = \Psi_t(x_0, y_0)$, one obtains

$$x(t) \times y(t) \;=\; \cos^2(t) x_0 \times y_0 - \sin^2(t) y_0 \times x_0 \;=\; x_0 \times y_0\,.$$

   Different orbits lie therefore in different oriented planes. So the orbit space is the sphere S^2.
   (b) For $d = 3$, the orbit space is (analogous to $d = 2$) the Grassmann manifold of oriented two-dimensional planes in $\mathbb{R}^4$. The plane containing the trajectory with initial condition $(x, y) \in T_1 S^3$ is spanned by its orthonormal basis (x, y). The mapping

$$\Phi : T_1 S^3 \to S^2 \times S^2 \quad , \quad (x, y) \mapsto (xy^*, y^*x)\,,$$

   in which we have identified points $(a, b, c, d)^\top \in \mathbb{R}^4$ with quaternions $x = \left(\begin{smallmatrix} a+b\imath & c+d\imath \\ -c+d\imath & a-b\imath \end{smallmatrix}\right) \in \mathbb{H}$, maps $T_1 S^3$ into $S^2 \times S^2$, because $\|mn\| = \|m\|\|n\|$ for $m, n \in \mathbb{H}$. Moreover, for $\big(x(t), y(t)\big) = \Psi_t(x_0, y_0)$,

$$\begin{aligned} x(t)y(t)^* &= \big(\cos(t)x_0 + \sin(t)y_0\big)\big(-\sin(t)x_0 + \cos(t)y_0\big)^* \\ &= \sin(t)\cos(t)(y_0 y_0^* - x_0 x_0^*) + \cos^2(t) x_0 y_0^* - \sin^2(t) y_0 x_0^* \;=\; x_0 y_0^*. \end{aligned}$$

   This is because $\|x_0\| = \|y_0\| = 1$, hence $x_0 x_0^* = y_0 y_0^* = \mathbb{1}$, and $x_0 y_0^* = -y_0 x_0^* \in \Im\mathbb{H}$, because the vectors are orthogonal, hence $\mathrm{tr}(x_0 y_0^*) = 0$. Analogously, $y^*(t)x(t) = y_0^* x_0$. So we can factorize the mapping Φ through the rotation Ψ.
   For a quaternion $z \in \mathbb{H}$ of norm $\|z\| = 1$ and the image $(u, v) := (xy^*, y^*x)$ of the mapping Φ, one has

$$z - uzv = 2(\langle z, x\rangle\, x + \langle z, y\rangle\, y)\,. \tag{H.6}$$

   Therefore, the plane spanned by x and y can be reconstructed from the image. Since

$$\Phi(y, x) = (yx^*, x^*y) = -(xy^*, y^*x) = -\Phi(x, y)\,,$$

we also obtain its orientation.
Hence Φ/S^1 is injective. Surjectivity also follows from (H.6), by choosing $(u, v) \in S^2 \times S^2$ arbitrarily.
Hence Φ/S^1 is a diffeomorphism.

(c) Since (11.3) regularizes the *direction* of the Laplace-Runge-Lenz vector, this vector itself can be regularized, too. The formula for A given in the hint follows therefore from the formula $\|\hat{A}\|^2 = 2\|\hat{L}\|^2\hat{H} + Z^2$ for $\hat{A}$, which was proved in (11.3.13).
Therefore, for $E < 0$, we conclude $\|A\| \le Z$.
For $\|A\| < Z$, the ellipse is not degenerate, and there are exactly two periodic orbits with the same value of A in Σ_E.
In contrast, the vectors with $\|A\| = Z$ correspond to collision orbits, namely those periodic orbits in Σ_E that coincide with their image under time reversal (Definition 11.3).
Therefore the orbit space consists of two discs that are glued along their boundaries, which is S^2.
Conversely, $\|A\| \ge Z$ if $E > 0$, and in this case, the orbit space is a cylinder $S^1 \times \mathbb{R}$. □

H.12 Chapter 12, Scattering Theory

Exercise 12.7 on page 282 (Asymptotics of Scattering in a Potential):

1. By reversibility of the flow, it suffices to show the existence of the Cesàro limit $p^+(x_0) = \lim_{T\to+\infty} \frac{1}{T}\int_0^T p(t, x_0)\,\mathrm{d}t$ for all initial data $x_0 \in P$.
   - For $E > 0$ and $x_0 \in s_E^+$, this Cesàro limit exists by Theorem 12.5.
   - If $x_0 \in b^+$, then $p^+(x_0) = \lim_{T\to+\infty} \frac{q(T,x_0)}{T} = 0$, because $\limsup_T \|q(T, x_0)\| < \infty$.
   - For $E > 0$, one has $\Sigma_E = s_E^+ \cup b_E^+$; whereas for $E < 0$, one has $\Sigma_E = b_E^+$. This leaves the case $E = 0$. In this case, $\|p\| = \sqrt{2V(q)} \le c\,\langle q\rangle^{-\varepsilon}$. If we had $\limsup_{T\to\infty} \frac{\|q(T,x_0)\|}{T} > 0$, i.e., if there existed $k > 0$ and an increasing sequence of times $t_n \to \infty$ with $\frac{\|q(t_n,x_0)\|}{t_n} \ge k$, then one would in particular have the inequality $\langle q(t_n, x_0)\rangle > \left(\frac{2c}{k}\right)^{1/\varepsilon}$ for all $n \ge n_0$. But then, $\|q(t_{n+1}, x_0)\| - \|q(t_n, x_0)\|$ would be smaller than

$$\int_{t_n}^{t_{n+1}} \|p(t, x_0)\|\,\mathrm{d}t < c\int_{t_n}^{t_{n+1}} \langle q(t, x_0)\rangle^{-\varepsilon}\,\mathrm{d}t < \frac{k}{2}(t_{n+1} - t_n)\,,$$

   and thus $\limsup_{m\to\infty} \frac{\|q(t_m,x_0)\|}{t_m} \le k/2$, so this is impossible.

2. We write briefly $\big(p(t), q(t)\big) := \big(p(t, x_0), q(t, x_0)\big)$ and let $L(t) := q(t) \wedge p(t)$. Since by hypothesis, there exists a potential W with $W(q) = \tilde{W}(\|q\|)$ such that $V - W$ is short range, W is automatically long range, too.
As we check the Cauchy condition for the limit of L, we find for $0 < t_1 \le t_2$

$$\|L(t_2) - L(t_1)\| \;=\; \Big\| \int_{t_1}^{t_2} q(t) \wedge \nabla V\big(q(t)\big)\, \mathrm{d}t \Big\|$$
$$= \left\| \int_{t_1}^{t_2} q(t) \wedge \nabla\big[V\big(q(t)\big) - W\big(q(t)\big)\big]\, \mathrm{d}t \right\| \;\le\; \int_{t_1}^{t_2} \langle q(t)\rangle^{-1-\varepsilon}\, \mathrm{d}t \;\le\; c\, t_1^{-\varepsilon},$$

analogous to (12.1.8). So the asymptotic angular momenta (12.1.10) exist. The applicability to the singular molecule potentials (12.1.11) follows from the short range property of the differences $\frac{Z_k}{\|q - s_k\|} - \frac{Z_k}{\|q\|} = \mathcal{O}(\|q\|^{-2})$. □

Exercise 12.10 on page 285 (Scattering Orbits):

1. For given $E > 0$ and $\ell \in \mathbb{R}$, there exists a minimum radius

$$r_{\min} = -\frac{Z}{2E} + \sqrt{\frac{Z^2}{4E^2} + \frac{\ell^2}{2E}} \ge 0\,.$$

For $r_2 > r_1 \ge r_{\min}$, there exist times $t_2 > t_1 \ge t_{\min}$ such that the orbit $t \to q(t) \in \mathbb{R}^2$ realizes these distances, i.e., $\|q(t_{\min})\| = r_{\min},\ \|q(t_i)\| = r_i$.
It follows that $t_2 - t_1 = \int_{r_1}^{r_2} \frac{r}{\sqrt{2r^2E + 2Zr - \ell^2}}\, \mathrm{d}r$, because for times $t > t_{\min}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\|q(t)\| = \frac{\langle \dot{q}(t), q(t)\rangle}{\|q(t)\|} = \sqrt{2\left(E + \frac{Z}{\|q(t)\|} - \frac{\ell^2}{2\|q(t)\|^2}\right)} > 0\,.$$

2. Let $t \mapsto \big(p(t), q(t)\big)$ be a solution to the Hamiltonian equations with $\lim_{t\to+\infty} q_1(t) = +\infty$. We consider the dynamics of the second coordinate on the extended phase space $R := \mathbb{R}_p \times \mathbb{R}_q \times \mathbb{R}_t$ (so we drop the subscript 2). This dynamics is described by the time-dependent Hamiltonian

$$h : R \to \mathbb{R} \quad , \quad h(p, q, t) := \tfrac{1}{2}\big(p^2 + q_1^2(t) q^2\big)\,.$$

Similar to Example 10.44, we rewrite this time-dependent harmonic oscillator in polar coordinates. To this end, we use the generating function

$$s(q, \varphi, t) := \tfrac{1}{2} q_1(t) q^2 \cot(\varphi).$$

We get the time-dependent canonical transformation

$$p = \frac{\partial s}{\partial q} \quad \text{, hence} \quad \varphi = \arctan\left(q_1(t)\frac{q}{p}\right)$$
$$J = -\frac{\partial s}{\partial \varphi} \quad \text{, hence} \quad J = \tfrac{1}{2}\left(q_1(t)q^2 + \frac{p^2}{q_1(t)}\right).$$

Thus $h(p, q, t) = q_1(t)J(p, q, t)$. The Hamiltonian k defined as

$$k(J, \varphi, t) = q_1(t)J + \frac{\partial s}{\partial t}\big(q(J, \varphi, t), \varphi, t\big)$$

generates the dynamics in the new coordinates. Explicitly, it is

$$k(J, \varphi, t) = q_1(t)J + \tfrac{1}{2}\frac{q_1'(t)}{q_1(t)} J \sin(2\varphi).$$

So we get $\dot{J} = -\frac{q_1'(t)}{q_1(t)} J \cos(2\varphi)$ and $\dot{\varphi} = q_1(t) + \frac{1}{2}\frac{q_1'(t)}{q_1(t)} \sin(2\varphi)$.
Under our hypothesis that $\lim_{t\to+\infty} q_1(t) = +\infty$, the perturbation theory of Chapter 15.2 can be applied to this system of differential equations.
Accordingly, by Theorem 15.13, there exists $c > 0$ such that

$$|J(t) - J(t_0)| < \frac{c}{q_1(t_0)} \qquad \big(t - t_0 \in (0, q_1(t_0))\big).$$

Thus J is a so-called *adiabatic invariant* (see Arnol'd [Ar2, Chapter 10E]). In the limit of large t_0, it follows from the energy bound $E \geq h\big(p(t), q(t), t\big) = q_1(t)J\big(p(t), q(t), t\big)$ and $q_1(t) \to +\infty$ that $J\big(p(t_0), q(t_0), t_0\big) = 0$.
But then $p(t_0) = q(t_0) = 0$, and hence in the original coordinates, $p_2(t) = q_2(t) = 0$ for all times $t \in \mathbb{R}$. □

Exercise 12.13 on page 291 (Møller Operators in 1D):
As the Møller operators conserve the energies according to (12.2.2), one has

$$H^{(0)} \circ \mathcal{S}(x) = H^{(0)} \circ (\Omega^+)^{-1} \circ \Omega^-(x) = H \circ \Omega^-(x) = H^{(0)}(x).$$

This implies for $x = (p, q) \in P_+^{(0)}$ and $x' = (p', q') := \mathcal{S}(x)$ that $|p'| = |p|$.

But for $E > V_{\max}$, the energy shell $\Sigma_E = H^{-1}(E)$ consists of exactly two connected components, for which $\text{sign}(p)$ is $+1$ and -1 respectively. Since they are invariant under the flow Φ_t, we infer $p' = p$.

Since $V_{\max} \geq 0$ and we assumed $E > V_{\max}$, the Lebesgue integral defining $\tau(p)$ exists for short range potentials V. But for arbitrary intervals $I := [q_1, q_2] \subset \mathbb{R}$, the time the free particle spends in I equals $\frac{q_2 - q_1}{\sqrt{2E}}$, whereas the time spent there by the particle in the potential equals $\int_{q_1}^{q_2} \big(2(E - V(q))\big)^{-1/2} \mathrm{d}q$; so the formula $\tau(p) = \int_{\mathbb{R}} \left[(2\big(E - V(q)\big)^{-1/2} - (2E)^{-1/2}\right] \mathrm{d}q$ follows. □

Exercise 12.24 on page 298 (Scattering at High Energies):

(a) As $E > \|V\|_\infty$, the direction $\theta(t) := \frac{p(t)}{\|p(t)\|}$ is defined for all times t. The rate of change of the direction,

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{-\nabla V\big(q(t)\big)\|p(t)\| + \Big\langle \nabla V\big(q(t)\big), \frac{p(t)}{\|p(t)\|}\Big\rangle p(t)}{\|p(t)\|^2},$$

has a norm that is bounded by $\left\|\frac{\mathrm{d}\theta}{\mathrm{d}t}\right\| \le \frac{\|\nabla V\|_\infty}{\|p(t)\|} \le \frac{\|\nabla V\|_\infty}{v_{\min}}$, because $\|p(t)\| \in [v_{\min}, v_{\max}]$.

(b) For all times $t > 0$, one has

$$\|q(t) - q(0)\| = \left\|\int_0^t p(s)\,\mathrm{d}s\right\| \le v_{\max}\, t\,.$$

A lower bound is obtained by projecting the orbit onto its initial direction. For times $t > 0$, one gets

$$\begin{aligned}\|q(t) - q(0)\| &\ge \Big\langle q(t) - q(0)\,,\ \frac{p(0)}{\|p(0)\|}\Big\rangle = \int_0^t \Big\langle p(s)\,,\ \frac{p(0)}{\|p(0)\|}\Big\rangle\,\mathrm{d}s \\ &\ge \int_0^t \Big\langle p(0) - \int_0^s \nabla V(q(\tau))\,d\tau\,,\ \frac{p(0)}{\|p(0)\|}\Big\rangle\,\mathrm{d}s \\ &\ge \|p(0)\|\,t - \tfrac{1}{2}\|\nabla V\|_\infty\, t^2 \ge t\,(v_{\min} - \tfrac{1}{2}\|\nabla V\|_\infty t)\,.\end{aligned}$$

This bound is positive when $t < \frac{2v_{\min}}{\|\nabla V\|_\infty}$, and it is maximal for $t = \frac{v_{\min}}{\|\nabla V\|_\infty}$.

(c) Now assume $t_0 \in \mathbb{R}$ is a time when $\langle p(t_0), q(t_0)\rangle = 0$. (If such a time t_0 should fail to exist, one instead considers a sequence $(t_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty}\langle p(t_n), q(t_n)\rangle = 0$.)
The lower bound for E implies $v_{\min}^2 \ge 4R\|\nabla V\|_\infty$.
From (b), we get for $t_\pm := t_0 \pm \frac{2R}{v_{\min}}$ that $\left|\Big\langle q(t_\pm) - q_0, \frac{p_0}{\|p(0)\|}\Big\rangle\right| \ge R$; so the orbit leaves the ball of radius R after time t_+ and before time t_-, and it will continue as a straight line outside this ball.
According to (a), the change in angle in the interval $[t_-, t_+]$ is at most

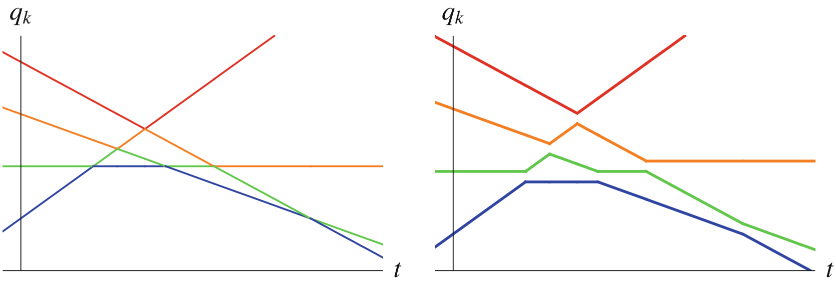
$$\int_{t_-}^{t_+} \left\|\frac{\mathrm{d}\theta}{\mathrm{d}t}\right\|\mathrm{d}t \le \frac{4R\,\|\nabla V\|_\infty}{v_{\min}^2} = \frac{2R\,\|\nabla V\|_\infty}{E - \|V\|_\infty} \le 1\,. \qquad \square$$

Exercise 12.35 on page 308 (Billiard Balls):

(a) Denoting the positions of the centers of the balls in 1 dimension as $q_i \in \mathbb{R}$ ($i = 1, \ldots, n$), we may assume that they are in increasing order for all times t. If the radii of the balls are R, and $q_N := \frac{\sum_{k=1}^n m_k q_k}{\sum_{k=1}^n m_k}$ denotes their center of mass, then the mapping

$$q_k \mapsto q_k - 2kR + \Delta \quad \text{with} \quad \Delta := 2R \frac{\sum_{\ell=1}^n \ell\, m_\ell}{\sum_{\ell=1}^n m_\ell} \qquad (k = 1, \ldots, n)$$

does not change the center of mass, but reduces the distances $q_{k+1} - q_k$ by $2R$. Thus the dynamics gets reduced to the dynamics of n point particles (with radius 0), see the figure.



Now if these particles have equal masses $m := m_1 = \ldots = m_n$, then their velocities $v_i = p_i/m_i$ after a collision of the i^{th} and the $(i+1)^{\text{st}}$ particle according to (12.5.1) satisfy

$$v_i^+ = v_{i+1}^- \quad \text{and} \quad v_{i+1}^+ = v_i^- \; .$$

In other words, the particles exchange their velocities.
So if one graphs the positions $q_1, \ldots, q_n$ as functions of time, one obtains n straight lines, with vertical intercepts q_i and slopes v_i, see the figure. As n straight lines intersect at most $\binom{n}{2}$ times, $\binom{n}{2}$ is the maximum number of collisions. This maximum number occurs exactly if all initial velocities are distinct.
In Newton's cradle, the masses are equal, too. Typically, the balls in their rest position are almost in contact. If one moves the first k balls to the left and releases them simultaneously, then the last k balls will end up moving to the right after all the collisions, which are perceived as *a single* collision. The way they are suspended, there is a restoring force, and the process repeats itself in reverse order.

(b) As can be seen from part (a), for a time t at which there is no collision, the solution $\big(p(t, x_0), q(t, x_0)\big)$ is continuous in the initial conditions x_0. Even multiple collisions can be extended continuously.
This is not the case when the masses are different. But even if we only consider initial conditions that avoid multi-particle collisions, the fact that the masses are different has an effect. For instance, one ball between two balls of much higher masses will be able to collide frequently.
Galperin has shown in [Galp] that for three balls with masses m_1, m_2, and m_3, and pairwise different initial velocities, the number of collisions is equal to

$$\left\lceil \frac{\pi}{\arccos \sqrt{\frac{m_1 m_3}{(m_1+m_2)(m_2+m_3)}}} \right\rceil$$

(where $\lceil \cdot \rceil$ denotes the *ceiling* function). So there are at least two collisions, for equal masses there are 3, and the number diverges if $m_2 \searrow 0$.

This result can be seen as follows: In the configuration space $\mathbb{R}^3_q$ of the three particles, one introduces the center of mass system, i.e., the plane

$$\tilde{E} := \left\{ q \in \mathbb{R}^3 \;\middle|\; \sum_{i=1}^3 m_i q_i = 0 \right\} .$$

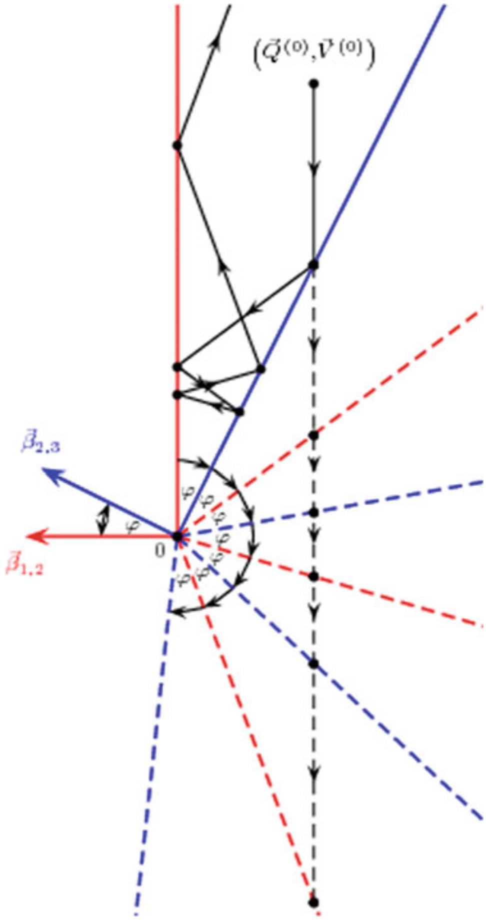
The linear mapping $Q := Mq$ with $M := \mathrm{diag}\big(\sqrt{m_1}, \sqrt{m_2}, \sqrt{m_3}\big)$ maps this plane to the plane $E := M\tilde{E} = \left\{ Q \in \mathbb{R}^3 \mid \sum_{i=1}^3 \sqrt{m_i}\, Q_i = 0 \right\}$. Collisions correspond to the straight lines

$$\left\{ Q \in E \mid Q_i/\sqrt{m_i} = Q_{i+1}/\sqrt{m_{i+1}} \right\}$$

$(i = 1, 2)$. The benefit from this change of coordinates is that, in a collision of the i^{th} with the $(i+1)^{\text{st}}$ particle, the vector of transformed velocities $V := Mv$ gets reflected in the line $\mathrm{span}\big(e_i/\sqrt{m_i} - e_{i+1}/\sqrt{m_{i+1}}\big) \subset E$. As the angle φ between these two lines is of the form

$$\varphi = \arccos\left(\sqrt{\frac{m_1 m_3}{(m_1+m_2)(m_2+m_3)}} \right) ,$$

the claimed formula is obtained by unfolding the configuration space in the plane E as seen in the figure.[34] □



[34] Image: courtesy of Markus Stepan.

Exercise 12.37 on page 308 (Potential of a Centrally Symmetric Mass Distribution):

(a) Both the Lebesgue measure on $\mathbb{R}^3$ and the ball B_R^3 are invariant under the action of $O \in \mathrm{SO}(3)$. Therefore, letting $y := O^{-1}x$, one has $V(Oq) =$

$$\int_{B_R^3} \frac{\rho(x)}{\|Oq - x\|}\,\mathrm{d}x = \int_{B_R^3} \frac{\rho(O^{-1}x)}{\|O(q - O^{-1}x)\|}\,\mathrm{d}x = \int_{B_R^3} \frac{\rho(y)}{\|O(q-y)\|}\,\mathrm{d}y = V(q)\,.$$

(b) $\|q - x\| = \sqrt{x_1^2 + x_2^2 + (x_3 - a)^2}$. Using spherical coordinates

$$x_1 = r\cos(\theta)\cos(\varphi) \quad , \quad x_2 = r\cos(\theta)\sin(\varphi) \quad , \quad x_3 = r\sin(\theta)\,,$$

this quantity equals $\sqrt{r^2 + a^2 - 2ra\sin(\theta)}$.

(c) The Jacobi determinant occurring under the integral when transitioning to spherical coordinates is $r^2\cos(\theta)$. Thus, for $a = \|q\| > 0$,

$$V(q) = 2\pi \int_0^R \int_{-\pi/2}^{\pi/2} \frac{\tilde{\rho}(r) r^2 \cos\theta}{\sqrt{r^2 + a^2 - 2ra\sin\theta}}\,\mathrm{d}\theta\,\mathrm{d}r$$
$$\frac{2\pi}{a} \int_0^R \int_{-ar}^{ar} \frac{\tilde{\rho}(r) r\,\mathrm{d}u}{\sqrt{r^2 + a^2 - 2u}}\,\mathrm{d}r = \frac{2\pi}{a} \int_0^R r\big(|a + r| - |a - r|\big)\tilde{\rho}(r)\,\mathrm{d}r\,.$$

(d) For $a > R$, one has therefore $V(q) = \frac{2\pi}{a}\int_0^R 2r^2\tilde{\rho}(r)\,\mathrm{d}r$.
In spherical coordinates, the integral that gives the total mass or charge is $M = 2\pi \int_0^R \int_{-\pi/2}^{\pi/2} \tilde{\rho}(r) r^2 \cos\theta\,\mathrm{d}\theta\,\mathrm{d}r = 2\pi \int_0^R 2r^2\tilde{\rho}(r)\,\mathrm{d}r$.

(e) For $a \in (0, R]$, one gets $V(q) = \frac{2\pi}{a}\int_0^a 2r^2\tilde{\rho}(r)\,\mathrm{d}r + \frac{2\pi}{a}\int_a^R 2ra\tilde{\rho}(r)\,\mathrm{d}r$.
Hence for $\tilde{V}(\|q\|) = V(q)$, one obtains $\frac{2}{a}\partial_a \tilde{V}(a) = -\frac{4\pi}{a^3}\int_0^a 2r^2\tilde{\rho}(r)\,\mathrm{d}r$ and $\partial_a^2 \tilde{V}(a) = +\frac{4\pi}{a^3}\int_0^a 2r^2\tilde{\rho}(r)\,\mathrm{d}r - 4\pi\tilde{\rho}(a)$, thus altogether $-\Delta V = 4\pi\rho$.

(f) We use the constant of gravitation $G \approx 6.67 \cdot 10^{-11}\,\frac{\mathrm{m}^3}{\mathrm{kg\,s}^2}$. Since by Newton's law for the force, the orbit period is $T = \sqrt{\frac{3\pi}{\rho G}}$, it only depends on the average density ρ of the body, not on its radius. This density doesn't vary much ($\rho_{\mathrm{Earth}} \approx 5\,500\,\frac{\mathrm{kg}}{\mathrm{m}^3}$, the earth being the densest planet, $\rho_{\mathrm{Sun}} \approx 1\,400\,\frac{\mathrm{kg}}{\mathrm{m}^3}$). So the orbit around the sun would take two and three quarter hours.[35] □

Exercise 12.45 on page 313 (Cluster Projections):
For a cluster decomposition $\mathcal{C} = \{C_1, \ldots, C_k\} \in \mathcal{P}(N)$, the linear mappings

$$\Pi_{\mathcal{C}}^E : M \to M \quad , \quad \Pi_{\mathcal{C}}^E(q)_\ell = \frac{1}{m_{C_i}} \sum_{j \in C_i} m_j q_j$$

[35] In fact (as pointed out by Victor Weisskopf in his 'Modern physics from an elementary point of view') its order of magnitude can be calculated as a combination of constants of nature like the Bohr radius.

(with particle index $\ell \in C_i$) onto the Euclidean vector space $(M, \langle\cdot,\cdot\rangle_{\mathcal{M}})$ are orthogonal projections:

$$\left(\Pi^E_{\mathcal{C}} \circ \Pi^E_{\mathcal{C}}(q)\right)_\ell = \frac{1}{m_{C_i}} \sum_{j\in C_i} \frac{m_j}{m_{C_i}} \sum_{r\in C_i} m_r q_r = \frac{1}{m_{C_i}} \sum_{r\in C_i} m_r q_r = \Pi^E_{\mathcal{C}}(q)_\ell$$

and

$$\left\langle \Pi^E_{\mathcal{C}}(q'), q\right\rangle_{\mathcal{M}} = \sum_{\ell=1}^{n} m_\ell \left\langle \Pi^E_{\mathcal{C}}(q')_\ell, q_\ell\right\rangle = \sum_{i=1}^{k} \sum_{\ell\in C_i} m_\ell \left\langle \tfrac{1}{m_{C_i}} \textstyle\sum_{r\in C_i} m_r q'_r, q_\ell\right\rangle$$

$$= \sum_{i=1}^{k} \sum_{r\in C_i} m_r \left\langle q'_r, \frac{1}{m_{C_i}} \sum_{\ell\in C_i} m_\ell q_\ell \right\rangle = \sum_{i=1}^{k} \sum_{r\in C_i} m_r \left\langle q'_r, \Pi^E_{\mathcal{C}}(q)_r\right\rangle$$
$$= \left\langle q', \Pi^E_{\mathcal{C}}(q)\right\rangle_{\mathcal{M}} .$$

Therefore, $\Pi^I_{\mathcal{C}} = 1\!\!1_M - \Pi^E_{\mathcal{C}}$ is an orthogonal projection as well:

$$\Pi^I_{\mathcal{C}} \circ \Pi^I_{\mathcal{C}} = 1\!\!1_M - 1\!\!1_M \circ \Pi^E_{\mathcal{C}} - \Pi^E_{\mathcal{C}} \circ 1\!\!1_M + \Pi^E_{\mathcal{C}} \circ \Pi^E_{\mathcal{C}} = 1\!\!1_M - \Pi^E_{\mathcal{C}} = \Pi^I_{\mathcal{C}}$$

and $\left(\Pi^I_{\mathcal{C}}\right)^* = 1\!\!1_M - (\Pi^E_{\mathcal{C}})^* = 1\!\!1_M - \Pi^E_{\mathcal{C}} = \Pi^I_{\mathcal{C}}$. □

Exercise 12.47 on page 315 (Moments of Inertia):
By hypothesis, the partition $\mathcal{D} \in \mathcal{P}(N)$ is finer than $\mathcal{C} \in \mathcal{P}(N)$. Therefore, the center-of-mass projections satisfy $\Pi^E_{\mathcal{C}}\,\Pi^E_{\mathcal{D}} = \Pi^E_{\mathcal{D}}\,\Pi^E_{\mathcal{C}} = \Pi^E_{\mathcal{D}}$, thus we can compare the external moments of inertia $J^E_{\mathcal{C}} = J \circ \Pi^E_{\mathcal{C}}$ and $J^E_{\mathcal{D}} = J \circ \Pi^E_{\mathcal{D}}$ by $J^E_{\mathcal{D}} = J^E_{\mathcal{C}} \circ \Pi^E_{\mathcal{D}}$. Since J is convex, and thus $J^E_{\mathcal{C}}$ is convex as well, it follows that $J^E_{\mathcal{D}} \leq J^E_{\mathcal{C}}$. □

H.13 Chapter 13, Integrable Systems and Symmetries

Exercise 13.12 on page 343 (Action of the Planar Pendulum):
We use the complete elliptic integrals of the first kind, namely[36]

$$K(k) \;=\; \int_0^{\pi/2}\left(1 - k^2 \sin^2(\varphi)\right)^{-1/2} \mathrm{d}\varphi \;=\; \int_0^1 \frac{\mathrm{d}x}{\sqrt{(1-x^2)(1-k^2x^2)}}\,,$$

and of the second kind:

$$E(k) \;=\; \int_0^{\pi/2}\left(1 - k^2 \sin^2(\varphi)\right)^{1/2} \mathrm{d}\varphi \;=\; \int_0^1 \sqrt{\tfrac{1-k^2x^2}{1-x^2}}\,\mathrm{d}x\,.$$

The action $I(h) = \int_{H^{-1}([-1,h])} \mathrm{d}p_\psi\, \mathrm{d}\psi$ is calculated as follows:

[36] When using them, be aware that frequently their definition varies among authors!

- For $h > 1$ with $\varphi := \psi/2$, it is

$$I(h) = 4\int_0^{\pi} \sqrt{2(h+\cos(\psi))}\,\mathrm{d}\psi = 8\int_0^{\pi/2} \sqrt{2(h+1-2\sin^2\varphi)}\,\mathrm{d}\varphi$$
$$= 8\sqrt{2(h+1)}E\left(\sqrt{\tfrac{2}{h+1}}\right).$$

- For $h \in (-1, 1)$, we use the substitution $x := \sqrt{\frac{1-\cos(\psi)}{1+h}}$.
For $\psi \in [0, \arccos(-h)]$, one has therefore $x \in [0, 1]$. With the abbreviation $k := \sqrt{\frac{1+h}{2}}$, the integrand of $I(h) = 4\int_0^{\arccos(-h)} \sqrt{2(h+\cos(\psi))}\,\mathrm{d}\psi$ becomes $\sqrt{2(h+\cos(\psi))} = 2k\sqrt{1-x^2}$, and $\mathrm{d}\psi = \frac{2k}{\sqrt{1-k^2x^2}}\,\mathrm{d}x$. Therefore

$$I(h) = 16k^2\int_0^1 \sqrt{\frac{1-x^2}{1-k^2x^2}}\,\mathrm{d}x = 16\big(E(k) - (1-k^2)K(k)\big).$$

Thus the derivative of the continuous and increasing function $I : [-1, \infty) \to [0, \infty)$ exists everywhere except at $h = 1$, and it is equal to the orbit period. At $h = 1$, it diverges because of the upper equilibrium. In the limit $h \nearrow \infty$, we see that $I(h)$ is asymptotic to $4\pi\sqrt{2h}$. □

Exercise 13.29 on page 359 (Coadjoint Orbits):

(a) Using $\mathcal{O}(\xi) = \{g\xi g^{-1} \mid g \in G\}$ and parametrizing a neighborhood of $e \in G$ by $g = \exp(u)$ with $g^{-1} = \exp(-u)$, the claim $T_\xi\mathcal{O}(\xi) = \{[u, \xi] \mid u \in \mathfrak{g} \subset \mathrm{Mat}(n, \mathbb{R})\}$ follows from the product rule for the derivative.
(b) The symplectic form on the left hand side of (13.5.14), defined by the reduction, is Ad_g^*-invariant. Since $\mathrm{Ad}_g[u, v] = [\mathrm{Ad}_g u, \mathrm{Ad}_g v]$, the 2-form on the right hand side of (13.5.14) is Ad_g^*-invariant as well. The proof that the two sides are identical can be found e.g. in Chapter 14 of Marsden and Ratiu [MR].
(c) By formula (13.5.13) for the momentum mapping $J : T^*G \to \mathfrak{g}^*$ and the defining equation $X_\xi(e) = \xi$ of the infinitesimal generator $X_\xi : G \to TG$, the restriction of J to $T_e^*G = \mathfrak{g}^*$ is the identity. □

Exercise 13.32 on page 360 (Hamiltonian S^1-Action on S^2):
On the cylinder $\mathcal{Z} := \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 = 1, |x_3| < 1\}$ with the area form as its symplectic form, $H_{\mathcal{Z}} : \mathcal{Z} \to \mathbb{R}$, $x \mapsto x_3$ generates the S^1-action $\Phi_t(x) = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix} x$. By Exercise 10.30, the radial projection $F : \mathcal{Z} \to S^2$ is a symplectomorphism onto its image, and $H \circ F = H_{\mathcal{Z}}$. It is only the poles $\begin{pmatrix} 0 \\ 0 \\ \pm 1 \end{pmatrix}$, where H is extremal, that are not hit by F. Hence H, too, generates an S^1-action. □

Exercise 13.37 on page 364 (Ky-Fan Maximum Principle for Hermitian Matrices): As $A \in \mathrm{Herm}(d, \mathbb{C})$ has eigenspaces for the eigenvalues λ_i that are mutually orthogonal, we see, by choosing orthonormal eigenvectors $x_1, \ldots, x_k \in \mathbb{C}^d$ for the eigenvalues $\lambda_1, \ldots, \lambda_k$, that at least the inequality

$$\sum_{i=1}^{k} \lambda_i \;\leq\; \max_{(x_1,\ldots,x_k)\in V_k(\mathbb{C}^d)} \sum_{i=1}^{k} \langle x_i, Ax_i \rangle$$

holds. Conversely, given $(x_1, \ldots, x_k) \in V_k(\mathbb{C}^d)$, we extend these to an orthonormal basis $(x_1, \ldots, x_d) \in V_d(\mathbb{C}^d)$. Then the representation of A with respect to this basis will have certain diagonal values $b_{i,i}$, which, by the Theorem of Schur and Horn, can be written as convex combinations of the $\lambda_{\sigma(i)}$ with $(\sigma \in S_d)$:

$$b_{i,i} = \sum_{\sigma\in S_d} c_\sigma \lambda_{\sigma(i)} \quad \text{with} \quad c_\sigma \geq 0 \quad , \quad \sum_{\sigma\in S_d} c_\sigma = 1\,.$$

The maximum of $\sum_{i=1}^{k} b_{i,i}$ will be taken on at a vertex of the polytope $\Pi\big(\Lambda^{-1}(\lambda)\big)$. But due to the weak ordering $\lambda_1 \geq \lambda_2 \geq \ldots$ of the eigenvalues, one such vertex giving a maximum is $\sigma = \mathrm{Id}$. □

H.14 Chapter 14, Rigid and Non-Rigid Bodies

Exercise 14.2 on page 367: $(v, O)^{-1} = (-O^{-1}v, O^{-1})$. □

Exercise 14.8 on page 373 (Pseudoforces):

1. In the dimension-independent formula (14.2.5), one has for $d = 2$ the matrix $B = O^{-1}\,O' = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix} \begin{pmatrix} -\sin(\varphi) & -\cos(\varphi) \\ \cos(\varphi) & -\sin(\varphi) \end{pmatrix} \varphi' = \varphi' \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \varphi' \mathbb{J}$.
   This implies $B^2 = -(\varphi')^2\,\mathbb{1}$, $B' = \varphi''\mathbb{J}$, and thus the claim.
2. This formula is immediate from (14.2.5) and $Bv = i(\omega)v = \omega \times v$ (see (13.4.8)).
3. The absolute value $2m\|C'\|\,\|\omega\|$ of the Coriolis force does not depend on the geographical location, since the direction of travel is orthogonal to the axis of rotation of the earth. With $\|\omega\| = 2\pi/T$, $T = 86\,400$ s and $\|C'\| = 20\,000/3\,600$ m/s, the force is about 0.08 Newton. For comparison, a bar of chocolate weighs about 1 N.
   The vertical component of the Coriolis force points downward, the horizontal component points north; so it adds to the weight of the biker and pulls him to the right. This can be seen by the right-hand rule for the vector product in the formula $-2m\omega \times C'$ for the Coriolis force.
   Given the geographic latitude of about 53.5° for Berlin, the components are divided up as about 0.048 N down and 0.064 N right.
   The absolute value of the force component pulling to the right is independent of the direction of travel. □

Exercise 14.10 on page 375 (Steiner's Theorem):

- From the formula for $I(q)$, the minimality of $I\big(T_{q_N}(q)\big)$ follows directly, because the term $m_N\|q_N\|^2(1\!\!1_3 - P_N)$ in the sum is positive semidefinite and vanishes exactly if the center of mass is $q_N = 0$.
- As for the formula itself, it follows from $I = J\,1\!\!1_3 - \tilde{I}$, with $\tilde{I}$ from (14.2.3), because $J(q) - J\big(T_{q_N}(q)\big) = \sum_{k=1}^{n} 2m_k(\langle q_N, q_k\rangle - \|q_N\|^2) = m_N\|q_N\|^2$. Analogously, one finds $\tilde{I}_0(q) - \tilde{I}_{q_N(q)}(q) = m_N\|q_N\|^2 P_N$. □

Exercise 14.11 on page 376 (Principal Moments of Inertia):

- It suffices to consider the case $n = 3$, because for $q_4 := \dots q_n := 0$, the principal moments of inertia are determined by the first three mass points. Now let $q_k = e_k$. Then $I = \mathrm{diag}(m_2 + m_3, m_1 + m_3, m_1 + m_2)$. For masses $0 \le m_3 \le m_2 \le m_1$, one has therefore $(I_1, I_2, I_3) = (m_2 + m_3, m_1 + m_3, m_1 + m_2)$. The inequalities of the masses translate into $I_1 \le I_2 \le I_3 \le I_1 + I_2$.
- Analogously, one can show, in an orthonormal basis in which I is diagonal, that the inequality $I_3 \le I_1 + I_2$ is valid in general for arbitrary $n \in \mathbb{N}$.
- If the center of mass for the three mass points is the origin of $\mathbb{R}^3$, then $\mathrm{span}(q_1, q_2, q_3)$ is at most 2-dimensional. In this case, $I_1 + I_2 = I_3$, so the principal moments of inertia are convex combinations of those of a disc and of a rod.
- It follows from the symmetry under arbitrary rotations that a homogeneous ball with center 0 has three equal principal moments of inertia.
- For a rod that is oriented in 1-direction, one has $I = \mathrm{diag}(0, I_2, I_2)$.
- The tensor of inertia I for a circular distribution of mass is obtained by integrating $I = \int_0^{\pi/2} I(\varphi)\,\mathrm{d}\varphi$ for the tensors of inertia $I(\varphi)$ of a configuration of equal masses at the four points $\pm e_1(\varphi)$ and $\pm e_2(\varphi)$ with

$$e_1(\varphi) := \cos(\varphi)e_1 + \sin(\varphi)e_2 \quad \text{and} \quad e_2(\varphi) := \cos(\varphi)e_2 - \sin(\varphi)e_1.$$

  As one has $I(\varphi) = I(0)$ in this case, the claim $I = \mathrm{diag}(I_1, I_1, 2I_1)$ follows from the fact that the projection onto the 1-2 plane leaves all the four points $\pm e_1$, $\pm e_2$ invariant, whereas the projections onto the 1-3 and the 2-3 planes only leave two of them invariant, mapping the other two into zero. □

Exercise 14.17 on page 383 (Fast Top):

The initial condition $\beta(0) = 0$ (hence $u(0) = 1$) requires $\ell_z = \ell_Z$. The initial condition $\dot{\beta}(0) = 0$ means that $E = V_{\mathrm{eff}}(0) = \frac{\ell_Z^2}{2I_3} + 1$. Altogether, this implies

$$U_{\mathrm{eff}}(u) = I_1^{-1}(1-u)^2\Big[2(1+u) - \tfrac{\ell_Z^2}{I_1}\Big].$$

The rotation is therefore stable for $\frac{\ell_Z^2}{I_1} > 4$, i.e., for frequency $|\omega| > 2/\sqrt{I_1}$. □

Exercise 14.20 on page 386 (Euclidean Symmetries):

1. The action by $(v, O) \in \mathbb{SE}(d)$ on $\mathbb{R}^d$ is given by $x \mapsto Ox + v$, see (14.1.2). Here, $v \in \mathbb{R}^d$ and $O \in \mathrm{SO}(d)$. Now if $K \subset \mathbb{R}^d \times \mathbb{R}^d$ is compact, and therefore bounded by some bound $b > 0$, then the point $(x, Ox + v)$ can only lie in K if $\|v\| \le 2b$, because $\|Ox + v\| \ge \|v\| - \|Ox\| = \|v\| - \|x\|$. Since the rotation group $\mathrm{SO}(d)$ is compact, the pre-image of K under the mapping $\big((O, v), x)\big) \mapsto (x, Ox + v)$ is therefore also compact.
   The diagonal action of $\mathbb{SE}(d)$ on $\mathbb{R}^{nd}$ can be viewed as the action of a closed subgroup of $\mathbb{SE}(nd)$. It is therefore proper by the same argument. Since the open and dense subset Q is invariant under this action, the claim also follows for the action restricted to Q.
2. The action of $G := \mathbb{R}^+ \times \mathbb{SE}(2)$ on $\mathbb{R}^2$ given by $x \mapsto \lambda O(x + v)$ is an action of the Lie group G. In complex notation, G acts on $\mathbb{R}^2 \cong \mathbb{C}$ as the multiplication by a complex number $z = \lambda O \in \mathbb{C}^*$ and translation by $v \in \mathbb{C}$.
   G also acts diagonally on the configuration space $(\mathbb{R}^2)^3 \cong \mathbb{C}_q^3$ for three particles in the plane. If we set $z := 1/(q_2 - q_1)$ and $v := -q_1$, then q_1 will be mapped to $0 \in \mathbb{C}$ and q_2 to $1 \in \mathbb{C}$. The image $w \in \mathbb{C} \setminus \{0, 1\}$ of q_3 then parametrizes the group orbits, and $\mathbb{C} \setminus \{0, 1\}$ is the image of the form sphere under stereographic projection. □

H.15 Chapter 15, Perturbation Theory

Exercise 15.5 on page 396 (Conditionally Periodic Motion):
The conditionally periodic motion on $\mathbb{T}^2$ with the frequency vector $(1, 1/12)$ has period 12. Its closed orbits hit the diagonal $\{(x, x) \in \mathbb{T}^2 \mid x \in S^1\}$ eleven times. Hence the hour and the minute hand meet 22 times during a day. □

Exercise 15.12 on page 400 (Virial Theorem for Space Averages):

(a) $\langle \{f, H\} \rangle_E = \int_{\Sigma_E} \{f, H\}\, \mathrm{d}\lambda_E = \int_{\Sigma_E} \frac{\mathrm{d}}{\mathrm{d}t} f \circ \Phi_t|_{t=0}\, \mathrm{d}\lambda_E = 0$, because the measure λ_E is Φ-invariant.
(b) Obvious.
(c) For $f(p, q) := \langle p, q \rangle$, the Poisson bracket

$$\{f, H\}(p, q) = \sum_{j=1}^{N} \Big(\tfrac{1}{2} \|p_j\|^2 - \big\langle q_j, \nabla U_j(q_j) \big\rangle - \sum_{k=j+1}^{N} \big\langle q_j, \nabla W_{j,k}(q_j - q_k) \big\rangle \Big)$$

has average 0. The first term in the sum is the mean kinetic energy.
(d) By compactness of the manifold with boundary G, there exists $\varepsilon > 0$ for which the function $q \mapsto U(q) = \mathrm{dist}(q, G)^2$ is smooth on $G_\varepsilon := \{q \in \mathbb{R}^3 \mid \mathrm{dist}(q, G) \le \varepsilon\}$. U equals ε^2 on ∂G_ε. Therefore, for parameter values $\lambda > E/\varepsilon^2$, the energy shell $\Sigma_{E,\lambda}$ is also smooth, because it projects with respect to all par-

ticle indices j onto their respective domain of space $G_{\sqrt{E/\lambda}}$. As $\lambda \nearrow \infty$, the volume $V_{E/\lambda}$ of $G_{\sqrt{E/\lambda}}$ decreases to the volume V of G.
The proof of the equation for the ideal gas is quite lengthy, so we only give a sketch here.

1. If we normalize the Liouville measure on $\Sigma_{E,\lambda}$ to be a probability measure $\mu_{E,\lambda}$ and integrate it over the velocities, we obtain the density
$$v(E,\lambda)^{-1}\big(2(1-\lambda\tilde{U}(\tilde{q})/E)_+\big)^{d/2-1}\, d\tilde{q} \quad , \text{with} \quad v(E,\lambda) \in (V, V_{E/\lambda}) \, ,$$
$\tilde{q} = (q_1, \ldots, q_N)$, and $\tilde{U}(\tilde{q}) = \sum_{k=1}^{N} U(q_k)$.
Therefore, in the limit $\lambda \nearrow \infty$, the integral of this density over G^N increases to 1.
2. Since $U \geq 0$, the kinetic energy T on $\Sigma_{E,\lambda}$ is smaller than E. But as $U\restriction_G = 0$, it follows that $\lim_{\lambda\nearrow\infty} \langle T \rangle_{E,\lambda} = E$. Therefore it follows from the virial theorem that $\lim_{\lambda\nearrow\infty} \langle M \rangle_{E,\lambda} = E$, where $M_\lambda(\tilde{q}) := \sum_{k=1}^{N} \langle q_k, \nabla\lambda U(q_k)\rangle$.
3. If G denotes the cube $[-\ell,\ell]^3$, and e_1, e_2, e_3 are orthogonal basis vectors of $\mathbb{R}^3$, then it follows that the sum of the projections onto these vectors is 1. The total pressure for instance on the lateral areas $\{\pm\ell\} \times [-\ell,\ell]^2$ can be defined as $\lim_{\lambda\nearrow\infty} \langle R_1 \rangle_{E,\lambda}$, with
$$R_i(\tilde{q}) := \sum_{k=1}^{N} \operatorname{sign}(q_k) \, \langle e_i, \lambda\nabla U(q_k)\rangle \, .$$
By the fundamental theorem, this integral becomes $2PV$ in the limit $\lambda \nearrow \infty$. Summing over the three basis vectors, we obtain $PV = \frac{2}{3}\langle T \rangle_E$.
4. For arbitrary smoothly bounded $G \subset \mathbb{R}^3$, the analogous statement follows using the Stokes theorem. □

Exercise 15.22 on page 411 (Relativistic Advance of the Perihelion):
The Hamilton function[37] (15.3.18) is of the form $H_\varepsilon = H_0 + \varepsilon K$, with the Hamilton function $H_0(p,q) = \frac{1}{2}\|p\|^2 - \frac{Z}{\|q\|}$ of the Kepler motion and $K(p,q) = -1/\|q\|^3$.

We denote the maximal flow generated by H_ε as Φ_ε. The Runge-Lenz vector A is invariant under the Kepler flow Φ_0. Therefore, within one period T of the Φ_ε-orbit $t \mapsto x(t)$, according to the first order of Hamiltonian perturbation theory, A changes by
$$\int_0^T \frac{dA}{dt}\, dt = \varepsilon\int_0^T \big[-D_p A\big(x(t)\big)\nabla_q K\big(x(t)\big) + D_q A\big(x(t)\big)\nabla_p K\big(x(t)\big)\big]\, dt + \mathcal{O}(\varepsilon^2).$$

The second term under the integral vanishes, the first yields, in view of
$$D_p A(x) = \begin{pmatrix} -q_2 p_2 & 2q_1 p_2 - q_2 p_1 \\ 2q_2 p_1 - q_1 p_2 & -q_1 p_1 \end{pmatrix} \quad \text{and} \quad \nabla_q K(x) = -\frac{3}{\|q\|^5}\begin{pmatrix} q_1 \\ q_2 \end{pmatrix},$$

[37] We omit the hats from (15.3.18) for simplicity.

the portion proportional to ε in $\int_0^T \frac{\mathrm{d}A}{\mathrm{d}t}\mathrm{d}t = 3\varepsilon\ell \int_0^T \left(\begin{smallmatrix} q_2 \\ -q_1 \end{smallmatrix}\right) / \|q\|^5 \, \mathrm{d}t + \mathcal{O}(\varepsilon^2)$. Let us use the angle φ, rather than time t, as the parameter. Using the conic section equation (1.7) for the Kepler ellipse with $\varphi_0 := 0$ and $\dot\varphi = \ell/r^2$, one has

$$3\varepsilon\ell \int_0^T \frac{\left(\begin{smallmatrix} q_2 \\ -q_1 \end{smallmatrix}\right)}{\|q\|^5}\,\mathrm{d}t = \frac{3\varepsilon Z^2}{\ell^4}\int_0^{2\pi} \bigl(1 + e\cos(\varphi)\bigr)^2 \left(\begin{smallmatrix} \sin\varphi \\ \cos\varphi \end{smallmatrix}\right) \mathrm{d}\varphi$$
$$= \frac{6\pi\varepsilon Z^2}{\ell^4} e \left(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right) = \frac{6\pi\varepsilon Z}{\ell^4}\|A\| \left(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right).$$

In the last step, we used $\|A\| = Ze$ from Exercise 11.22 (b). This perturbation is orthogonal to the vector A (Exercise 11.22 (a)). Therefore, its norm remains constant up to an error of order ε^2. Dividing by $\|A\|$ yields the change in the argument of A. □

Exercise 15.38 on page 430 (Nondegeneracy Conditions):

1. $\omega_1(I) = 1 + 2I_1, \omega_2(I) = 1$. Therefore, $\mathrm{D}\omega = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $\begin{pmatrix} \mathrm{D}\omega & \omega \\ \omega^\top & 0 \end{pmatrix} = \left(\begin{smallmatrix} 2 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{smallmatrix}\right)$.
2. $\omega_1(I) = 1 + 2I_1, \omega_2(I) = 1 - 2I_2$. $\mathrm{D}\omega = \left(\begin{smallmatrix} 2 & 0 \\ 0 & -2 \end{smallmatrix}\right)$ and $\begin{pmatrix} \mathrm{D}\omega & \omega \\ \omega^\top & 0 \end{pmatrix} = \left(\begin{smallmatrix} 2 & 0 & 1 \\ 0 & -2 & 1 \\ 1 & 1 & 0 \end{smallmatrix}\right)$. □

Exercise 15.47 on page 434 (Continued Fraction Expansion):

- The existence of $\lim_{n\to\infty}\omega_n$ follows from parts 1 and 3 of Theorem 15.46, because these state that the subsequences with even resp. odd index are increasing resp. decreasing, and that the difference of the elements of the sequence converges to 0.
- The fact that the sequence converges to ω will be a consequence of the relation

$$\omega = \frac{p_n + h^{(n)}(\{\omega\})\, p_{n-1}}{q_n + h^{(n)}(\{\omega\})\, q_{n-1}} \qquad (n \in \mathbb{N}_0). \tag{H.15.1}$$

(H.15.1) follows by induction: $\frac{p_0 + h^{(0)}(\{\omega\})\, p_{-1}}{q_0 + h^{(0)}(\{\omega\})\, q_{-1}} = \frac{\lfloor\omega\rfloor + \{\omega\}}{1+0} = \omega$ and, letting $\rho := h^{(n)}(\{\omega\})$, one has

$$\frac{p_{n+1} + h(\rho)\, p_n}{q_{n+1} + h(\rho)\, q_n} = \frac{a_{n+1}p_n + p_{n-1} + h(\rho)\, p_n}{a_{n+1}q_n + q_{n-1} + h(\rho)\, q_n} = \frac{(\lfloor 1/\rho\rfloor + h(\rho))p_n + p_{n-1}}{\lfloor 1/\rho\rfloor + h(\rho))q_n + q_{n-1}}$$
$$= \frac{\rho^{-1}p_n + p_{n-1}}{\rho^{-1}q_n + q_{n-1}} = \frac{p_n + \rho p_{n-1}}{q_n + \rho q_{n-1}} = \omega.$$

- Now (H.15.1) says, in view of the positivity of $h^{(n)}(\{\omega\})$, that ω lies between $\omega_n = \frac{p_n}{q_n}$ and $\omega_{n-1} = \frac{p_{n-1}}{q_{n-1}}$. See also *Theorem 14* in KHINCHIN [Kh]. □

H.16 Chapter 16, Relativistic Mechanics

Exercise 16.6 on page 448 (Lorentz Boosts):

1. By definition, the Lorentz boosts are the positive matrices in the restricted Lorentz group $\mathrm{SO}^+(3,1)$.
   Firstly, the given $L(v)$ are of this type. This is obvious for $L(0)$.
   For $v \in \mathbb{R}^3$ with $0 < \|v\| < 1$, one can check $L(v)^\top I L(v) = I$ by calculation. The positivity of $L(v)$ then follows, with a vector $w = \begin{pmatrix} \tilde{w} \\ w_4 \end{pmatrix} \in \mathbb{R}^4$, from the identity
   $$\langle w, L(v)w\rangle = \gamma(v)(\langle \tilde{w}, v\rangle + v_4)^2 + \|\tilde{w}\|^2 \geq 0\,.$$
   Conversely, let $A = \begin{pmatrix} a & b \\ c^\top & d \end{pmatrix} \in \mathrm{SO}^+(3,1)$ (with $a \in \mathrm{Mat}(3,\mathbb{R})$, $b, c \in \mathbb{R}^3$ and $d \in \mathbb{R}$) be positive and $A \neq 1\!\!1_4$. The symmetry of A implies $b = c$ and $a^\top = a$, and the positivity implies $d > 0$.
   The condition $A^\top I A = I$ contains the relation $d^2 - \|b\|^2 = 1$, the eigenvalue equation $ab = db$, and $a^2 = 1\!\!1_3 + b \otimes b^\top$. Therefore $d \geq 1$. The case $d = 1$ cannot occur, because in this case, b would have to be 0 and thus $a = 1\!\!1_3$, which however contradicts the hypothesis $A \neq 1\!\!1_4$. Therefore, $d = \gamma$, with $\gamma := \sqrt{1 + \|b\|^2} > 1$ and $v := b/\gamma$, satisfies $\|v\| \in (0,1)$. One also has the relation $\gamma = (1 - \|v\|^2)^{-1/2}$. The matrix $\tilde{a} := 1\!\!1_3 - P_v + \gamma(v) P_v$ is positive, and its square satisfies $\tilde{a}^2 = 1\!\!1_3 - P_v + \gamma(v)^2 P_v = 1\!\!1_3 + b \otimes b^\top$.
   We have thus shown that $A = L(v)$.
2. The claimed relation $d^2 = 1 + \|b\|^2$ is shown as in the first part of the exercise. We already know that P is positive. It first needs to be shown that $\tilde{O}$ is orthogonal. To this end, one exploits the equation $AIA^\top = I$ (which is valid since along with A, its transpose $A^\top$ is also in $\mathrm{O}(3,1)$):
   $$aa^\top - b \otimes b^\top = 1\!\!1_3 \quad , \quad ac = bd \quad , \quad c^\top c = d^2 - 1\,.$$
   We obtain $\tilde{O} = O \oplus 1$ with $O = a[1\!\!1_3 - P_v + \gamma P_v] - bc^\top$. Calculation of $OO^\top$ yields $1\!\!1_3$. From $[1\!\!1_3 - P_v + \gamma P_v]c = \gamma c = dc$ and $ac = bd$, one gets $Oc = (d^2 - c^\top c)b = b$.
3. In the equation $P = L(c/d)$, one can read off c and d from the product $A = \begin{pmatrix} a & b \\ c^\top & d \end{pmatrix} := L(v_1)L(v_2)$. Namely one has $d = \gamma(v_1)\gamma(v_2)(1 + \langle v_1, v_2\rangle)$ and
   $$c = [(1\!\!1_3 - P_{v_1}) \;+\; \gamma(v_1) P_{v_1}]\gamma(v_2) v_2 \;+\; \gamma(v_1)\gamma(v_2) v_1\,.$$
   Therefore, $u = c/d$ equals
   $$\frac{[(1\!\!1_3 - P_{v_1})/\gamma(v_1) + P_{v_1}]v_2 + v_1}{1 + \langle v_1, v_2\rangle} = \frac{v_1 + v_2 + \left(1 - \sqrt{1 - \|v_1\|^2}\right)\left(v_1 \langle v_1, v_2\rangle / \|v_1\|^2 - v_2\right)}{1 + \langle v_1, v_2\rangle}\,.$$
   The identity $v_1 \times (v_1 \times v_2) = v_1 \langle v_1, v_2\rangle - v_2 \|v_1\|^2$ implies (16.2.6).

The square of $u = \left(v_1 + v_2 + \frac{v_1 \times (v_1 \times v_2)}{1+\sqrt{1-\|v_1\|^2}}\right) / (1 + \langle v_1, v_2 \rangle)$ has the denominator

$$\|v_1 + v_2\|^2 + 2\frac{\langle v_2, v_1 \times (v_1 \times v_2)\rangle}{1 + \sqrt{1 - \|v_1\|^2}} + \left\| \frac{v_1 \times (v_1 \times v_2)}{1 + \sqrt{1 - \|v_1\|^2}} \right\|^2$$

$$= \|v_1 + v_2\|^2 - \|v_1 \times v_2\|^2 \, \frac{2(1 + \sqrt{1 - \|v_1\|^2}) - \|v_1\|^2}{(1 + \sqrt{1 - \|v_1\|^2})^2} .$$

The last factor is 1. Because of $\|v_1 \times v_2\|^2 + \langle v_1, v_2\rangle^2 = \|v_1\|^2 \|v_2\|^2$, one further obtains

$$\|u\|^2 = \frac{\|v_1 + v_2\|^2 - \|v_1 \times v_2\|^2}{(1 + \langle v_1, v_2\rangle)^2} = 1 - \frac{\left(1 - \|v_1\|^2\right)\left(1 - \|v_2\|^2\right)}{(1 + \langle v_1, v_2\rangle)^2} < 1 .$$

4. From the relation $D_{12} = D_{21}^\top$ and equation (16.2.7), we conclude that $L(u_{21}) = \tilde{D}_{12}\, L(u_{12})\, \tilde{D}_{12}^\top$, hence it follows by (16.2.8) that $u_{21} = D_{12} u_{12}$. Since u_{21} and u_{12} lie in the subspace spanned by v_1 and v_2, the rotation axis of D_{12} is orthogonal to this space. □

Exercise 16.8 on page 449 (Minkowski Product):
By means of a Lorentz transformation, we can achieve that $v = (0, 0, 0, v_4)^\top$, and thus $\langle v, w\rangle_{3,1} = -v_4 w_4$.

On the other hand, a spacelike vector can be Lorentz transformed into the form $v = (v_1, 0, 0, 0)^\top \neq 0$, and therefore it has the $\langle \cdot, \cdot \rangle_{3,1}$-orthogonal spacelike subspace $\mathrm{span}(e_2, e_3)$. □

Exercise 16.14 on page 454 (Modified Twin Paradox):
In the formulas from Example 16.13, the time difference is

$$t_W - t_O = \frac{u_E}{c}\left(2\frac{v_E}{c} + \mathcal{O}\left(\frac{v_{\max}^3}{c^3}\right)\right).$$

This age difference of the two snails after their tour around the earth is, in the limit of vanishing speed, equal to $2\frac{u_E v_E}{c^2} \approx 0.41\mu$s (microseconds). □

Exercise 16.22 on page 465 (Galilei Group):

(a) The Galilei transformation is composed of a translation

$$(p_1, \ldots, p_n, E; q_1, \ldots, q_n t) \mapsto (p_1, \ldots, p_n, E; q_1 + \Delta q, \ldots, q_n + \Delta q, t + \Delta t), \tag{H.16.1}$$

a rotation

$$(p_1, \ldots, p_n, E; q_1, \ldots, q_n, t) \mapsto (O p_1, \ldots, O p_n, E; O q_1, \ldots, O q_n, t) , \tag{H.16.2}$$

and a boost

$$(p_1, \ldots, p_n, E; q_1, \ldots, q_n, t) \mapsto \tag{H.16.3}$$
$$\big(p_1 + m_1 v, \ldots, p_n + m_n v, E + \langle v, p_N\rangle + \tfrac{1}{2} m\|v\|^2; q_1 + vt, \ldots, q_n + vt, t\big)$$

of the extended phase space P_n.
The translation has derivative $\mathbb{1}$ and is therefore symplectic. As for the phase space rotation, it follows just as in Example 13.17 that it is symplectic.
The derivative of the boost is of the form $\left(\begin{smallmatrix} A & B \\ C & D\end{smallmatrix}\right)$ with submatrices A, B, C, $D \in \mathrm{Mat}(nd+1, \mathbb{R})$, $B = C = 0$, and $A = D^\top = \begin{pmatrix} \mathbb{1}_{nd} & 0 \\ v \ldots v & 1 \end{pmatrix}$.
Since we are using the symplectic form (16.5.5) with the configuration space $M = \mathbb{R}^{nd}_q$, modifying the test from Exercise 6.26 (b) with the time reversal I yields that the boost is symplectic as well.

(b) Since the potential depends only on the distances $q_k - q_\ell$, the Hamiltonian H is invariant under the translations (H.16.1). The requirement $V_{k,\ell}(Oq) = V_{k,\ell}(q)$ of rotation invariance implies invariance with respect to (H.16.2). The boost (H.16.3) does not change H, because

$$m = m_1 + \ldots + m_n \quad \text{and} \quad p_N = p_1 + \ldots + p_n\,. \qquad \square$$

Exercise 16.25 on page 467 (Constant Acceleration):
The acceleration $g > 0$ causes a spacelike distance of $1/g$ of the observer from the point of intersection of the lines. For $g = 10\ \mathrm{m/s}^2$, the distance is therefore $c^2/g \approx 9 \cdot 10^{12}$ km, which is approximately one light year. $\square$

H.17 Chapter 17, Symplectic Topology

Exercise 17.13 on page 480 (Elliptic Billiard):

- The orbits $\{(\pm a_1, 0; \mp 1, 0)\}$ and $\{(0, \pm a_2; 0, \mp 1)\}$ on the axes have period 2.
- An orbit of period 2 must belong to a billiard trajectory that is a segment that is orthogonal to $T_{p_i}\mathcal{C}$ at both of its end points $p_1, p_2 \in \mathcal{C}$. Therefore, $T_{p_1}\mathcal{C}$ is parallel to $T_{p_2}\mathcal{C}$, and hence $p_1 = -p_2$; so the segment passes through the center of the ellipse. Since $a_1 < a_2$, the segment is a semiaxis. $\square$

H.18 Appendices

Exercise A.47 on page 501 (Differential Topology):

1. The derivative $f'(t) = 3t^2$ of $f : \mathbb{R} \to \mathbb{R},\ t \mapsto t^3$ vanishes only at $t = 0$. So f is injective, and it is only at 0 that it fails to be immersive. Since $\lim_{t\to\pm\infty} f(t) = \pm\infty$, it follows that $f(\mathbb{R}) = \mathbb{R}$.

2. For $f : \mathbb{R} \to \mathbb{R}^2,\ t \mapsto \binom{t^3}{t^2}$, one has $f'(t) = \binom{3t^2}{2t}$. Therefore $\lim_{\pm t \searrow 0} \frac{f'(t)}{\|f'(t)\|} = \binom{0}{\pm 1}$, and $f(\mathbb{R})$ is not a submanifold.
3. For $k \in \mathbb{R}$, the derivatives of the rose curves $f_k : \mathbb{R} \to \mathbb{R}^2,\ t \mapsto \cos(kt)\binom{\sin t}{\cos t}$ are equal to $f_k'(t) = -k\sin(kt)\binom{\sin t}{\cos t} + \cos(kt)\binom{\cos t}{-\sin t}$. So they have norm $\sqrt{\cos^2(kt) + k^2\sin^2(kt)} > 0$, and therefore the f_k are immersions. For $k \neq 0$, one has $0 \in f_k(\mathbb{R})$, but in general, there are different directions $f_k'(t)$ for the different choices of t for which $f_k(t) = 0$. If that happens, $f_k(\mathbb{R})$ is not a submanifold.
4. f is 2π-periodic, therefore not injective. But we conclude that $S^1 \subset \mathbb{R}^2$ is a submanifold because $S^1 = g^{-1}(1)$ for $g(x) := \|x\|^2$.
5. $f : \mathbb{R} \to \mathbb{R}^2,\ t \mapsto \exp(-t^2)\binom{t}{t^3}$ has the derivative $f'(t) = \binom{1-2t^2}{t^2(3-2t^2)}e^{-t^2} \neq 0$. Now f is an odd function, and for $t > 0$, the function $t \mapsto \frac{f_2(t)}{f_1(t)} = t^2$ is strictly increasing. Therefore, f is injective. On the other hand, $\lim_{t\to\pm\infty} f(t) = 0 = f(0)$, so f is merely an injective immersion, but not an imbedding.
6. For $f \in C^1(M, N)$ and $m \in M,\ \operatorname{rank}(T_m f) \leq \min\big(\dim(M), \dim(N)\big)$.
7. If $\pi : E \to B$ is a C^1 fiber bundle, then by Definition F.1, the mapping π is a C^1-mapping of manifolds. If $b \in B$, then there exists a neighborhood $U \subset B$ of b and a diffeomorphism $\Phi : \pi^{-1}(U) \to U \times F$ with $\Phi\big(\pi^{-1}(b')\big) = \{b'\} \times F$. Therefore for all $f \in F : \pi \circ \Phi^{-1}\big((b', f)\big) = b'$. This implies the submersion property of π. □

Exercise B.9 on page 508 (Volume Form):
If we expand the exterior product $\omega^{\wedge n} \in \mathbf{\Omega}^{2n}(\mathbb{R}^{2n})$ distributively, there are exactly $n!$ nonvanishing terms in the sum. To bring these into standard order with respect to the dual basis $\alpha_1, \ldots, \alpha_{2n}$, namely into the form $\bigwedge_{i=1}^{n}(\alpha_i \wedge \alpha_{i+n})$, an even number of transpositions is needed in each case. In this process, the sign does not change. On the other hand, $\bigwedge_{i=1}^{n}(\alpha_i \wedge \alpha_{i+n}) = (-1)^{\binom{n}{2}} \bigwedge_{j=1}^{2n} \alpha_j$, because successively commuting $\alpha_{2n-1}, \alpha_{2n-2}, \ldots, \alpha_{n+1}$ to the $(2n-1)^{\text{st}}, (2n-2)^{\text{nd}}, \ldots, (n+1)^{\text{st}}$ position requires $1, 2, \ldots, n-1$ transpositions. □

Exercise B.16 on page 510 (Invariance of the Volume Form):
By Theorem E.5.5, $f^*(\omega^{\wedge n}) = (f^*\omega)^{\wedge n} = \omega^{\wedge n}$. So because of Exercise B.9, the volume form remains invariant. □

Exercise B.24 on page 515 (Pull-back of Exterior Forms):
The required formula for the pull-back of differential forms follows from the corresponding statement for exterior forms, namely Theorem E.5.5. □

Exercise B.32 on page 521 ($\wedge$-antiderivation):
The claim that the inner product is a $\wedge$-antiderivation follows by restriction to the tangential spaces $T_m M$, namely from an analogous formula for exterior forms.
For exterior forms, the formula can be checked on a basis of the Grassmann algebra $\mathbf{\Omega}^*(E)$ over the vector space E. In this case, it follows from the antisymmetry of the exterior product of 1-forms. □

Exercise E.25 on page 555 (Exponential Mapping for GL$(n, \mathbb{R})$**):**
Left invariance of the vector field $X^{(\xi)} = X_L^{(\xi)} : G \to TG$ with $X^{(\xi)}(e) = \xi \in \mathfrak{g}$ means, in terms of the left action $L_h : G \to G,\ g \mapsto h \circ g \quad (h \in G)$, that

$$\big((L_h)_* X^{\xi}\big)(f) = X^{\xi}(f) \qquad (f \in G).$$

For a vector field $Y : G \to TG$, this *push-forward* is of the form $((L_h)_* Y)(f) = \mathrm{D}L_h(h^{-1} \circ f)\big(Y(h^{-1} \circ f)\big)$. In case $Y = X^{(\xi)}$, one has $Y\big(h^{-1} \circ f\big) = h^{-1} \circ f\,\xi$. Since for the group $G = \mathrm{GL}(n, \mathbb{R})$, the action of L_h means multiplication by h from the left, the claim $X_L^{(\xi)}(g) = g\,\xi$ follows. The formula for right invariant vector fields is proved analogously.
The flow $\Phi^{(\xi)} : \mathbb{R} \times G \to G$ generated by $X_L^{(\xi)}$ is of the form $\Phi_t^{(\xi)}(g) = g \exp(\xi t)$, because $\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t^{(\xi)}(g)|_{t=0} = X_L^{(\xi)}(g) = g\,\xi$. Therefore the commutator of the left invariant vector fields, applied to $f \in C^\infty(G, \mathbb{R})$, is given by

$$\begin{aligned}
[X^{(\xi)}, X^{(\eta)}]\, f(g) &= \lim_{\varepsilon \to 0} \varepsilon^{-2}\left[f\left(\Phi_\varepsilon^{(\xi)} \circ \Phi_\varepsilon^{(\eta)}(g)\right) - f\left(\Phi_\varepsilon^{(\eta)} \circ \Phi_\varepsilon^{(\xi)}(g)\right)\right] \\
&= \lim_{\varepsilon \to 0} \varepsilon^{-2}\big[f\big(\exp(\varepsilon\xi)\exp(\varepsilon\eta)g\big) - f\big(\exp(\varepsilon\eta)\exp(\varepsilon\xi)g\big)\big] \\
&= df(g)[\xi, \eta] = X^{([\xi,\eta])} f(g)\,.
\end{aligned}$$

□

Exercise E.27 on page 556 (Lie Groups and Lie Algebras):

1. **Unitary Group:** A curve $c \in C^1\big(I, \mathrm{U}(n)\big)$ has a tangent vector $\dot{c}(0) \in \mathrm{Alt}(n, \mathbb{C})$, because $c(s)^* = c(s)^{-1}$ implies $\dot{c}^*(0) = -\dot{c}(0)$. Conversely, if $X \in \mathrm{Alt}(n, \mathbb{C})$, then X and $X^* = -X$ commute, and therefore

$$\exp(X)\exp(X)^* = \exp(X)\exp(X^*) = \exp(X + X^*) = \mathbb{1}\,,$$

hence $\exp(X) \in \mathrm{U}(n)$. This shows that the Lie algebra is $\mathfrak{u}(n) = \mathrm{Alt}(n, \mathbb{C})$, and

$$\dim\big(\mathrm{U}(n)\big) = \dim_{\mathbb{R}}\big(\mathrm{Alt}(n, \mathbb{C})\big) = n^2.$$

Special Unitary Group: For a curve $c \in C^1\big(I, \mathrm{SU}(n)\big)$ in the subgroup $\mathrm{SU}(n)$ of $\mathrm{U}(n)$, one concludes in addition to the previous that the tangent vector $\dot{c}(0) \in \mathrm{Alt}(n, \mathbb{C})$ has trace 0 because $0 = \frac{\mathrm{d}}{\mathrm{d}t}\det(c(t))|_{t=0} = \mathrm{tr}(\dot{c}(0))$. Therefore the Lie algebra $\mathfrak{su}(n) = \{X \in \mathrm{Alt}(n, \mathbb{C}) \mid \mathrm{tr}(X) = 0\}$, and

$$\dim\big(\mathrm{SU}(n)\big) = \dim_{\mathbb{R}}\big(\mathfrak{su}(n)\big) = n^2 - 1\,.$$

Symplectic Group: The Lie algebra $\mathfrak{sp}(\mathbb{R}^{2n})$ of $\mathrm{Sp}(\mathbb{R}^{2n})$ was already determined in Exercise 6.26, and also its dimension $n(2n + 1)$.

2. (Isomorphisms of Lie algebras)

(a) With $i : \mathbb{R}^3 \to \mathfrak{so}(3)$ from (13.4.8) and $i(a)b = a \times b$, the Jacobi identity implies: $\big(i(a_1)i(a_2) - i(a_2)i(a_1)\big)b = a_1 \times (a_2 \times b) - a_2 \times (a_1 \times b) = a_1 \times (a_2 \times b) + a_2 \times (b \times a_1) = (a_1 \times a_2) \times b$.

(b) Using as a basis of $\mathfrak{su}(2)$ the *Pauli matrices* multiplied by $\sqrt{-1}$, namely the matrices $(\tau_1, \tau_2, \tau_3) := \left(\left(\begin{smallmatrix} 0 & \imath \\ \imath & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} \imath & 0 \\ 0 & -\imath \end{smallmatrix}\right)\right)$, the mapping

$$\mathfrak{so}(3) \longrightarrow \mathfrak{su}(2) \quad , \quad \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix} \longmapsto \tfrac{1}{2}\sum_{i=1}^{3} a_i \tau_i$$

is such an isomorphism, because $[\tau_i, \tau_j] = 2\sum_{k=1}^{3} \varepsilon_{ijk}\tau_k$.

(c) Restricting the given imbedding to the imaginary quaternions $\Im\,\mathbb{H}$,

$$bi + cj + dk \mapsto \begin{pmatrix} b\imath & c+d\imath \\ -c+d\imath & -b\imath \end{pmatrix} = b\tau_3 - c\tau_2 + d\tau_1 \ ,$$

ij gets mapped to $-\tau_3\tau_2 = \tau_1$, and jk to $-\tau_2\tau_1 = \tau_3$, and ki to $\tau_1\tau_3 = -\tau_2$. (Also ii, jj and kk get mapped to $\tau_1^2 = \tau_2^2 = \tau_3^2 = -\mathbb{1}$). □

Exercise E.31 on page 559 (Adjoint Representation):

Using the Lie algebra isomorphism $i : \mathbb{R}^3 \to \mathfrak{so}(3)$, $a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \mapsto \begin{pmatrix} 0 & a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}$ from (13.4.8), we have to check the formula

$$O\, i(a)\, O^{-1} = i(Oa) \qquad \big(O \in \mathrm{SO}(3),\ a \in \mathbb{R}^3\big).$$

For elements of the form $O = \begin{pmatrix} c & -s & 0 \\ s & c & 0 \\ 0 & 0 & 1 \end{pmatrix}$ with $c^2 + s^2 = 1$, both sides are equal to $\begin{pmatrix} 0 & -a_3 & ca_2+sa_1 \\ a_3 & 0 & -ca_1+sa_2 \\ -ca_2-sa_1 & ca_1-sa_2 & 0 \end{pmatrix}$. However, any rotation of SO(3) is conjugate to such a rotation about the third axis (with the same angle of rotation). □

Exercise E.34 on page 550 (Adjoint Action):

- For the Lie group $G := \mathrm{GL}(n, \mathbb{R})$ with Lie algebra $\mathfrak{g} := \mathrm{Mat}(n, \mathbb{R})$, we infer from

$$\mathrm{Ad}_g(\eta) = g\eta g^{-1} \qquad (\eta \in \mathfrak{g},\ g \in G)$$

that

$$\mathrm{Ad}_g([\xi, \eta]) = \mathrm{Ad}_g(\xi\eta - \eta\xi) = g\xi g^{-1}\, g\eta g^{-1} - g\eta g^{-1}\, g\xi g^{-1} = [\mathrm{Ad}_g(\xi), \mathrm{Ad}_g(\eta)].$$

- Accordingly, $\exp\big(\mathrm{Ad}_g(\xi)\big) = \exp(g\xi g^{-1}) = g\exp(\xi)g^{-1}$.
- The adjoint representation of $\xi \in \mathfrak{g}$, applied to $\eta \in \mathfrak{g}$, yields

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ad}_{\exp(t\xi)}\, \eta\, \Big|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}\exp(t\xi)\ \eta\ \exp(-t\xi)\Big|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\exp(t\xi)\Big|_{t=0}\eta \ + \ \eta\, \frac{\mathrm{d}}{\mathrm{d}t}\exp(-t\xi)\Big|_{t=0} = [\xi, \eta]\,. \end{aligned}$$ □

Exercise E.37 on page 561 (Lie Group Actions):

1. It is clear that $\Phi : \mathrm{SO}(n) \times \mathbb{R}^n \to \mathbb{R}^n,\ (O, m) \mapsto Om$ is an action of the Lie group $\mathrm{SO}(n)$. Since $\mathrm{SO}(n)$ is compact (see Example E.37.2), the action is proper. For $n > 1$, it is not free, because $\Phi(\mathrm{SO}(n), \{0\}) = \{0\}$. Now $B = \mathbb{R}^n/\mathrm{SO}(n) = [0, \infty)$, because the other orbits are spheres centered at 0, parametrized by their radius. Consequently, $\partial B = \{0\}$.
2. The mapping $\Phi : \mathbb{R} \times M \to M,\ (t, m) \mapsto \left(\begin{smallmatrix} e^t & 0 \\ 0 & e^{-t} \end{smallmatrix}\right) x$ on $M = \mathbb{R}^2 \backslash \{0\}$ is a Lie group action of $(\mathbb{R}, +)$, because it is the restriction of a linear dynamical system to an open, flow-invariant subset of $\mathbb{R}^2$.
   It is free, because there are no periodic orbits, the origin having been removed. But the action is not proper, because the pre-image of the compact set $\big\{(x, y) \in M \times M \mid \|x - e_1\| \le \frac{1}{2} \ge \|y - e_2\|\big\}$ under the mapping $\mathbb{R} \times M \to M \times M,\ (t, m) \mapsto \big(m, \Phi(t, m)\big)$ is not compact.
   The topological space $B = M/\mathbb{R}$ is not Hausdorff, because the images of the points e_1 and e_2 are distinct, but do not have disjoint neighborhoods. □

Exercise F.7 on page 565 (Triviality of Principal Bundles):

- If the principal bundle $\pi : E \to B$ with the Lie group G as a typical fiber has a section $s : B \to E$ (i.e., $\pi \circ s = \mathrm{Id}_B$), then, denoting the group operation as $\Psi : E \times G \to E$, the mapping

  $$\rho : B \times G \to E \quad , \quad \rho(b, g) := \Psi_g \circ s(b)$$

  is a homeomorphism. Indeed, ρ is by definition continuous and bijective, and the inverse mapping $\rho^{-1}(e) = \big(\pi(e)),\ g(s \circ \pi(e), e)\big) \quad (e \in E)$ is continuous.
- Conversely, if a trivialization $\Phi : E \to B \times G$ exists and we denote the neutral element of the group as $e \in G$, then the mapping $s : B \to E,\ s(b) = \Phi^{-1}(b, e)$ is a section. □

Bibliography

[AA] V.I. Arnol'd, A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, Reading, 1968)

[Aa] J. Aaronson, *An Introduction to Infinite Ergodic Theory*. Mathematical Surveys and Monographs (American Mathematical Society, 1997)

[AB] R. Arens, D. Babbitt, The geometry of relativistic n particle interactions. Pac. J. Math. **28**, 243–274 (1969)

[AbMa] A. Abbondandolo, S. Matveyev, Middle-dimensional squeezing and non-squeezing behavior of symplectomorphisms. arXiv:1105.2931 (2011)

[ACL] M. Audin, A. Cannas da Silva, E. Lerman, *Symplectic Geometry of Integrable Hamiltonian Systems*. Advanced Courses in Mathematics, CRM Barcelona (Birkhäuser, Basel, 2003)

[AF] I. Agricola, T. Friedrich, *Global Analysis: Differential Forms in Analysis, Geometry, and Physics*. Graduate Studies in Mathematics 52 (American Mathematical Society, 2002)

[AG] V.I. Arnol'd, A. Givental, Symplectic geometry, in *Dynamical Systems IV, Symplectic Geometry and its Applications*. Encyclopaedia of Mathematical Sciences 4, ed. by V. Arnold, S. Novikov (Springer, Berlin, 1990)

[AGV] V.I. Arnol'd, S.M. Gusein-Zade, A.N. Varchenko, *Singularities of Differentiable Maps*, vol. 1. Monographs in Mathematics (Birkhäuser, Basel, 1985)

[AlK] A. Albouy, V. Kaloshin, Finiteness of central configurations of five bodies in the plane. Ann. Math. **176**, 535–588 (2012)

[AM] R. Abraham, J.E. Marsden, *Foundations of Mechanics*, 2nd edn., fourth printing (Benjamin/Cummings, Reading, 1982)

[Am] H. Amann, *Ordinary Differential Equations: An Introduction to Nonlinear Analysis*, De Gruyter Studies in Mathematics 13 (De Gruyter, Berlin, 1990)

[AN] V. Arnol'd, S. Novikov (eds.), *Dynamical Systems VII: Integrable Systems. Nonholonomic Dynamical Systems*. Encyclopaedia of Mathematical Sciences 16 (Springer, Berlin, 1993)

[AR] D. Arnold, J. Rogness, Möbius transformations revealed. Not. AMS **55**, 1226–1231 (2008)

[Ar1] V.I. Arnol'd, *Ordinary Differential Equations* (The MIT Press, Cambridge, 1973)

[Ar2] V.I. Arnol'd, *Mathematical Methods of Classical Mechanics*, Graduate Texts in Mathematics 60 (Springer, Berlin, 1989)

[Ar3] V.I. Arnol'd, *Geometrical Methods in the Theory of Ordinary Differential Equations*. Grundlehren der mathematischen Wissenschaften 250 (Springer, Berlin, 1997)

[Ar4] V.I. Arnol'd, Characteristic class entering in quantization conditions. Funct. Anal. Appl. **1**, 1–13 (1967)

[Ar5] V.I. Arnol'd, First steps in symplectic topology. Russ. Math. Surv. **41**, 1–21 (1986)

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

[Ar6] V.I. Arnol'd, *Huygens and Barrow, Newton and Hooke: Pioneers in Mathematical Analysis and Catastrophe Theory from Evolvents to Quasicrystals* (Birkhäuser, Basel, 1990)

[Art] E. Artin, *Geometric Algebra* (Interscience Publishers, New York, 1957)

[At] M. Atiyah, Convexity and commuting Hamiltonians. Bull. Lond. Math. Soc. **14**, 1–15 (1982)

[AvK] J. Avron, O. Kenneth, Swimming in curved space or the Baron and the cat. New J. Phys. **8**, 68–82 (2006)

[Bala] V. Baladi, *Positive Transfer Operators and Decay of Correlation*. Advanced Series in Nonlinear Dynamics (World Scientific, Singapore, 2000)

[Ball] W. Ballmann, *Lectures on Spaces of Nonpositive Curvature* (Birkhäuser, Basel, 1995)

[Ban] V. Bangert, On the existence of closed geodesics on two-spheres. Int. J. Math. **4**, 1–10 (1993)

[Bau] H. Bauer, *Measure and Integration Theory*, De Gruyter Studies in Mathematics 26 (De Gruyter, Berlin, 2001)

[BBT] O. Babelon, D. Bernard, M. Talon, *Introduction to Classical Integrable Systems* (Cambridge University Press, Cambridge, 2003)

[Berg] M. Berger, *A Panoramic View of Riemannian Geometry* (Springer, Berlin, 2003)

[Bern] J. Bernoulli, Problema novum ad cujus solutionem Mathematici invitantur. Acta Eruditorum **18**, 269 (1696)

[Berr] M. Berry, Singularities in waves and rays, in *Les Houches Lecture Series Session XXXV*, ed. by R. Balian, M. Kléman, J-P. Poirier (Amsterdam, North-Holland, 1981), pp. 453–543

[BFK] D. Burago, S. Ferleger, A. Kononenko, Uniform estimates on the number of collisions in semidispersing billiards. Ann. Math. **147**, 695–708 (1998)

[BH] H. Broer, G. Huitema, A proof of the isoenergetic KAM-theorem from the 'ordinary' one. J. Differ. Equ. **90**, 52–60 (1991)

[Bi1] G.D. Birkhoff, Proof of the ergodic theorem. Proc. Natl. Acad. Sci. **17**, 656–660 (1931)

[Bi2] G.D. Birkhoff, An extension of Poincaré's last geometric theorem. Acta Math. **47**, 297–311 (1925)

[Bi3] G.D. Birkhoff, *Dynamical Systems* (American Mathematical Society, 1927)

[BJ] Th. Bröcker, K. Jänich, *Introduction to Differential Topology* (Cambridge University Press, 1982)

[BKMM] A. Bloch, P. Krishnaprasad, J. Marsden, R. Murray, Nonholonomic mechanical systems with symmetry. Arch. Rational Mech. Anal. **136**, 21–99 (1996)

[BMZ] A. Bloch, J. Marsden, D. Zenkov, Nonholonomic dynamics. Not. AMS **52**, 324–333 (2005)

[BoSu] A. Bobenko, Y. Suris, *Discrete Differential Geometry: Integrable Structure*. Graduate Studies in Mathematics 98 (American Mathematical Society, Providence, 2008)

[Bo1] R. Bott, On the iteration of closed geodesics and the Sturm intersection theory. Comm. Pure Appl. Math. **9**, 171–206 (1956)

[Bo2] R. Bott, Lectures on Morse theory, old and new. Bull. Am. Math. Soc. **7**, 331–358 (1982)

[Br] H. Bruns, Über die Integrale des Vielkörper–Problems. Acta Math. **XI**, 25–96 (1887)

[BrSe] H. Broer, M. Sevryuk, KAM Theory: quasi-periodicity in dynamical systems, in *Handbook of Dynamical Systems*, ed. by B. Hasselblatt, A. Katok (North-Holland, 2007)

[BrSi] F. Brini, S. Siboni, Estimates of correlation decay in auto/endomorphisms of the n-torus. Comput. Math. Appl. **42**, 941–951 (2001)

[BT] R. Bott, L. Tu, *Differential Forms in Algebraic Topology*, Graduate Texts in Mathematics 82 (Springer, New York, 1995)

[Bu] L.A. Bunimovich et al., *Dynamical Systems, Ergodic Theory and Applications*. Encyclopaedia of Mathematical Sciences 100 (Springer, Berlin, 2000)

[CB] R. Cushman, L. Bates, *Global Aspects of Classical Integrable Systems* (Birkhäuser, Basel, 1997)

[CDD] Y. Choquet-Bruhat, C. DeWitt-Morette, M. Dillard-Bleick, *Analysis, Manifolds and Physics, Part 1: Basics* (North-Holland, 1982)

[CFKS] H.L. Cycon, R.G. Froese, W. Kirsch, B. Simon, *Schrödinger Operators with Application to Quantum Mechanics and Global Geometry*, Texts and Monographs in Physics (Springer, Berlin, 1987)

[CFS] I.P. Cornfeld, S.U. Fomin, Ya.G. Sinai, *Ergodic Theory*. Grundlehren der mathematischen Wissenschaften 245 (Springer, Berlin, 1982)

[CJS] D. Currie, T. Jordan, E. Sudarshan, Relativistic invariance and Hamiltonian theories of interacting particles. Rev. Modern Phys. **35**, 350–375 (1963)

[CL] A. Correia, J. Laskar, Mercury's capture into the 3/2 spin-orbit resonance including the effect of core-mantle friction. Icarus **201**(1), 1–11 (2009)

[CM] A. Chenciner, R. Montgomery, A remarkable periodic solution of the three-body problem in the case of equal masses. Ann. Math. **152**, 881–901 (2000)

[CN] C.A.A. de Carvalhoa, H.M. Nussenzveig, Time delay. Phys. Rep. **364**, 83–174 (2002)

[Cop] N. Copernicus, *De Revolutionibus Orbium Coelestium (On the Revolutions of the Heavenly Spheres)* (Nürnberg, 1543)

[Cr] F. Croom, *Basic Concepts of Algebraic Topology*, Undergraduate Texts in Mathematics (Springer, New York, 1978)

[CSM] R. Carter, G. Segal, I. Macdonald, *Lectures on Lie Groups and Lie Algebras*. London Mathematical Society Student Texts 32 (Cambridge University Press, Cambridge, 1995)

[Cv] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, G. Vattay, N. Whelan, A. Wirzba, *Chaos: Classical and Quantum*

[CW] N. Calkin, H. Wilf, Recounting the rationals. Am. Math. Monthly **107**, 360–363 (2000)

[De] M. Denker, *Einführung in die Analysis dynamischer Systeme* (Springer, Berlin, 2004)

[DFN] B.A. Dubrovin, A.T. Fomenko, S.P. Novikov, *Modern Geometry—Methods and Applications, Volumes I–III* (Springer, Berlin, 1992)

[DG] J. Dereziński, C. Gérard, *Scattering Theory of Classical and Quantum N-Particle Systems*. Texts and Monographs in Physics (Springer, Berlin, 1997)

[DH] F. Diacu, P. Holmes, *Celestial Encounters: The Origins of Chaos and Stability* (Princeton University Press, Princeton, 1996)

[Do] R. Douady: Applications du théorème des tores invariants. Thèse de 3e cycle, Université Paris, Paris 1982

[Dui] J.J. Duistermaat, Symplectic Geometry, Course notes of the Spring School (2004)

[Dum] S. Dumas, *The KAM Story. A Friendly Introduction to the Content, History, and Significance of Classical Kalmogorov-Arnold-Moser Theory* (World Scientific, Singapore, 2014)

[Ei1] A. Einstein, Zur Elektrodynamik bewegter Körper. Annalen der Physik und Chemie **17**, 891–921 (1905)

[Ei2] A. Einstein, Die Grundlage der allgemeinen Relativitätstheorie. Annalen der Physik **49** (whole series: **354**), 770–822 (1916)

[EH] I. Ekeland, H. Hofer, Symplectic topology and Hamiltonian dynamics. Mathematische Zeitschrift **100**, 355–378 (1989)

[El] J. Elstrodt, *Maß- und Integrationstheorie* (Springer, Berlin, 2009)

[Ep] M. Epple, *Die Entstehung der Knotentheorie: Kontexte und Konstruktionen einer modernen Mathematischen Theorie* (Vieweg, Braunschweig, 1999)

[Eu] L. Euler, De motu rectilineo trium corporum se mutuo attrahentium. Novi commentarii academiæ scientarum Petropolitanæ **11**, 144–151 (1767)

[Fei] M.J. Feigenbaum, The metric universal properties of period doubling bifurcations and the spectrum for a route to turbulence. Ann. New York. Acad. Sci. **357**, 330–336 (1980)

[Fej1] J. Féjoz, Démonstration du 'théorème d'Arnold' sur la stabilité du système planétaire (d'après Herman). Ergod. Theory Dyn. Syst. **24**, 1521–1582 (2004)

[Fej2] J. Féjoz, A proof of the invariant torus theorem of Kolmogorov. Regul. Chaotic Dyn. **17**, 1–5 (2012)

[FOF] J. Figueroa-O'Farrill, Deformations of the Galilean Algebra. J. Math. Phys. **30**, 2735–2739 (1989)

[FP] D. Fowler, P. Prusinkiewicz, Shell models in three dimensions, in *The Algorithmic Beauty of Sea Shells*, ed. by H. Meinhardt (Springer 4. Auflage, 2009)

[Gali] P. Galison, Einstein's Clocks, Poincaré's Maps: Empires of Time (W. W. Norton & Co, 2004)

[Galp] G. Galperin, Billiard balls count π, in *MASS Selecta* (AMS, Providence, RI, 2003), pp. 197–204

[Gal1] G. Galilei, *Dialogue Concerning the Two Chief World Systems* (Dialogo sopra i due massimi sistemi, Florenz 1632 English translation by Thomas Salusbury (1661)

[Gal2] G. Galilei, *Two New Sciences*. Discorsi e dimostrazioni matematiche (Leiden, 1638)

[Ge] H. Geiges, Contact geometry, in *Handbook of Differential Geometry*, vol. 2, ed. by F.J.E. Dillen, L.C.A. Verstraelen (North-Holland, Amsterdam, 2006)

[GHL] S. Gallot, D. Hulin, J. Lafontaine, *Riemannian Geometry* (Springer, Berlin, 1993)

[Gi] V.L. Ginzburg, The Weinstein conjecture and the theorems of nearby and almost existence, in *The Breadth of Symplectic and Poisson Geometry*. Festschrift in Honor of Alan Weinstein, ed. by J.E. Marsden and T.S. Ratiu (Birkhäuser, 2005), pp. 139–172

[GH] D.F. Griffiths, D.J. Higham, *Numerical Methods for Ordinary Differential Equations: Initial Value Problems*, Springer Undergraduate Mathematics Series (Springer, London, 2010)

[GiHi] M. Giaquinta, St. Hildebrandt, *Calculus of Variations II: The Hamiltonian Formalism*. Grundlehren der mathematischen Wissenschaften 311 (Springer, Berlin, 2004)

[GS1] V. Guillemin, S. Sternberg, *Symplectic Techniques in Physics* (Cambridge University Press, Cambridge, 1984)

[GS2] V. Guillemin, S. Sternberg, Convexity properties of the moment mapping. Inventiones mathematicae **67**, 491–513 (1982)

[GuHo] J. Guckenheimer, P. Holmes, *Nonlinear Oscillation, Dynamical Systems and Bifurcations of Vector Fields*, Applied Mathematical Sciences (Springer, Berlin, 2002)

[GZR] T. Geisel, A. Zacherl, G. Radons, Chaotic diffusion and $1/f$-noise of particles in two-dimensional solids. Z. Phys. B—Condens. Matter **71**, 117–127 (1988)

[HaMe] G. Hall, K. Meyer, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem* (Springer, Berlin, 1991)

[HaMo] M. Hampton, R. Moeckel, Finiteness of relative equilibria of the four-body problem. Inventiones Mathematicae **163**, 289–312 (2006)

[Hel] S. Helgason, *The Radon Transform*, 2nd edn. (Birkhäuser, Basel, 1999)

[Her] M. Herman, Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. Publ. Math. Inst. Hautes Étud. Sci. **49**, 5–233 (1979)

[Heu] H. Heuser, Gewöhnliche Differentialgleichungen (Teubner, 1995)

[Hil] S. Hildebrandt, *Analysis 1 and 2* (Springer, Berlin, 2002)

[Hirs] M. Hirsch, *Differential Topology*, Graduate Texts in Mathematics 33 (Springer, Berlin, 1988)

[Hirz] F. Hirzebruch, Division algebras and topology, in *Numbers*. Graduate Texts in Mathematics 123, ed. by H.-D. Ebbinghaus et al. (Springer, Berlin, Heidelberg, New York, 2008)

[HK] J. Hafele, R. Keating, Around the world atomic clocks: predicted relativistic time gains; observed relativistic time gains. Science **177**(166–168), 168–170 (1972)

[Hun] W. Hunziker, Scattering in classical mechanics, in *Scattering Theory in Mathematical Physics*, ed. by J.A. La Vita, J.-P. Marchand (Reidel, Dordrecht, 1974)

[Huy] Ch. Huygens, Discours de la cause de la pesanteur (Addition, 1690)

[HZ] H. Hofer, E. Zehnder, *Symplectic Invariants and Hamiltonian Dynamics* (Birkhäuser, Basel, 1994)

[Jac] C.G.J. Jacobi, Vorlesungen über Dynamik. Neunundzwanzigste Vorlesung. Œuvres complètes, tome **8**, 1–290 (1866)

[Jae] K. Jänich, *Topologie* (Springer, Berlin, Heidelberg, New York, 1987)

[JLJ] J. Jost, X. Li-Jost, *Calculus of Variations* (Cambridge University Press, Cambridge, 1999)

[Jo] J. Jost, *Riemannian Geometry and Geometric Analysis*. Universitext (Springer, Berlin, Heidelberg, New York, 2008)

[KB] U. Kraus, M. Borchers, Fast lichtschnell durch die Stadt. Physik in unserer Zeit, Heft **2**, 64–69 (2005)

[KH] A. Katok, B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge University Press, Cambridge, 1997)

[Kh] A. Ya, *Khinchin: Continued Fractions* (Dover, Mineola, 1997)

[Ki] A.A. Kirillov, Merits and demerits of the orbit method. Bull. Am. Math. Soc. **36**, 433–488 (1999)

[Kl] F. Klein, Vergleichende Betrachtungen über neuere geometrische Forschungen (Erlanger Programm) (1872)

[KL] V. Kaloshin, M. Levi, An example of Arnold diffusion for near-integrable Hamiltonians. Bull. Am. Math. Soc. **45**, 409–427 (2008)

[Kle] A. Klenke, *Probability Theory: A Comprehensive Course*, Universitext (Springer, Berlin, 2013)

[Kli1] W. Klingenberg, *A Course in Differential Geometry*, Graduate Texts in Mathematics 51 (Springer, Berlin, 1983)

[Kli2] W. Klingenberg, *Riemannian Geometry*. De Gruyter Studies in Mathematics 1, 2nd edn. (De Gruyter, Berlin, 1995)

[KMS] I. Kolář, P. Michor, J. Slovák, *Natural Operations in Differential Geometry* (Springer, Berlin, 1993)

[KN] S. Kobayashi, K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1996)

[Kn1] A. Knauf, Ergodic and topological properties of coulombic periodic potentials. Commun. Math. Phys. **110**, 89–112 (1987)

[Kn2] A. Knauf, Closed orbits and converse KAM theory. Nonlinearity **3**, 961–973 (1990)

[Kn3] A. Knauf, The n-centre problem of celestial mechanics for large energies. J. Eur. Math. Soc. **4**, 1–114 (2002)

[KP] Th Kappeler, J. Pöschel, *KdV & KAM*, Ergebnisse der Mathematik und ihrer Grenzgebiete 45D (Springer, Berlin, 2003)

[KR] M. Koecher, R., Remmert, Hamilton's quaternions, in *Numbers*. Graduate Texts in Mathematics 123, ed. by H.-D. Ebbinghaus et al. (Springer, Berlin, 2008)

[Kre] U. Krengel, *Ergodic Theorems*, De Gruyter Studies in Mathematics 6 (De Gruyter, Berlin, 1985)

[KS] A. Knauf, Ya. Sinai, *Classical Nonintegrability, Quantum Chaos*. DMV–Seminar 27 (Birkhäuser, Basel, 1997)

[KT] V.V. Kozlov, D.V. Treshchev, Billiards: a genetic introduction to the dynamics of systems with impacts. AMS Trans. Math. Monogr. **89** (1991)

[Lag] J. Lagrange, Essai sur le Probléme des trois Corps (1772). Œuvres de Lagrange/publ. par les soins de J.-A. Serret, Tome 6. (Gauthier-Villars, Paris, 1873)

[LaMe] A. Laub, K. Meyer, Canonical forms for symplectic and Hamiltonian matrices. Celest. Mech. Dyn. Astron. **89**, 213–238 (1974)

[Lap] P. de Laplace, *Essai philosophique sur les probabilités* (Courcier, 1814)

[Las1] J. Laskar, Large scale chaos and marginal stability in the solar system. Celest. Mech. Dyn. Astron. **64**, 115–162 (1996)

[Las2] J. Laskar, Existence of collisional trajectories of Mercury, Mars and Venus with the Earth. Nature **459**, 817–819 (2009)

[Leu] H. Leutwyler, A no-interaction theorem in classical relativistic Hamiltonian particle mechanics. Nuovo Cimento **37**, 556–567 (1965)

[Lev] M. Levi, *The Mathematical Mechanic: Using Physical Reasoning to Solve Problems* (Princeton University Press, Princeton, 2009)

[LF] L. Lusternik, A. Fet, Variational problems on closed manifolds. Dokl. Akad. Nauk. SSSR **81**, 17–18 (1951); Am. Math. Soc. Transl. **90** (1953)

[Li] P. Littelmann, Über Horns Vermutung, Geometrie, Kombinatorik und Darstellungstheorie. Jahresberichte der Deutschen Mathematiker-Vereinigung **110**, 75–99 (2008)

[LiMa] P. Libermann, Ch-M Marle, *Symplectic Geometry and Analytical Mechanics* (Reidel, Dordrecht, 1987)

[LL] A. Lichtenberg, M. Lieberman, Regular and chaotic dynamics, in *Applied Mathematical Sciences* (Springer, Berlin, 1993)

[Lon] Y. Long, Index theory for symplectic paths with applications, in *Progress in Mathematics 207* (Birkhäuser, Basel, 2002)

[Lou] A.K. Louis, *Inverse und schlecht gestellte Probleme* (Teubner, Stuttgart, 1989)

[LW] C. Liverani, M.P. Wojtkowski, *Ergodicity in Hamiltonian Systems*. Dynamics Reported IV (Springer, Berlin, 1995), pp. 130–202

[Mac] R.S. MacKay, I.C. Percival, Converse KAM: theory and practice. Commun. Math. Phys. **98**, 469–512 (1985)

[MaMe] L. Markus, K.R. Meyer, Generic Hamiltonian dynamical systems are neither integrable nor ergodic. Mem. AMS **144** (1974)

[Mat] J.N. Mather, Existence of quasiperiodic orbits for twist homeomorphisms of the annulus. Topology **21**, 457–467 (1982)

[McD] D. McDuff, Floer theory and low dimensional topology. Bull. Am. Math. Soc. **43**, 25–42 (2006)

[Me] B. Meißner: Die mechanische Wissenschaft und ihre Anwendung in der Antike, in *Physik/Mechanik*, ed. by A. Schürmann. Geschichte der Mathematik und der Naturwissenschaften 3 (Franz Steiner Verlag, 2005)

[Mil] J. Milnor, *Morse Theory*. Annals of Mathematics Studies 51. (Princeton University Press, Princeton, 1963)

[Min] H. Mineur, Réduction des systèmes mécaniques à *n* degrès de liberté admettant *n* intégrales premières uniformes en involution aux systèmes à variables séparées. J. Math. Pures Appl. **15**, 385–389 (1936)

[MKO] R.J. MacKay, R.W. Oldford, Scientific method, statistical method and the speed of light. Stat. Sci. **15**, 254–278 (2000)

[MM] R. MacKay, J. Meiss, Hamiltonian Dynamical Systems: A Reprint Selection (Taylor & Francis, 1987)

[MMC] J. Marsden, M. McCracken, *The Hopf Bifurcation and Its Applications*. Applied Mathematical Sciences 19. (Springer, New York, 1976)

[MMR] J. Marsden, R. Montgomery, T. Ratiu, *Reduction, Symmetry, and Phases in Mechanics*. Memoirs of the American Mathematical Society 436, vol. 88 (Providence, 1990)

[Mon1] R. Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics and Applications*. Mathematical Surveys and Monographs 91 (American Mathematical Society, 2002)

[Mon2] R. Montgomery, A new solution to the three-body problem. Not. AMS **48**, 471–481 (2001)

[Mos1] J. Moser, Is the solar system stable? Math. Intell. **1**, 65–71 (1978)

[Mos2] J. Moser, Various aspects of integrable Hamiltonian systems, in *Dynamical Systems* (C.I.M.E. Lect., Bressanone, 1978), Prog. Math. **8**, 233–290 (1980)

[Mos3] J. Moser, Regularization of Kepler's problem and the averaging method on a manifold. Commun. Pure Appl. Math. **23**, 609–636 (1970)

[Mos4] J. Moser, *Stable and Random Motion in Dynamical Systems* (Princeton University Press, Princeton, 1973)

[Mos5] J. Moser, Recent developments in the theory of Hamiltonian systems. SIAM Rev. **28**, 459–485 (1986)

[Mos6] J. Moser, Dynamical systems—past and present, in *Proceedings of the International Congress of Mathematicians*, Documenta Mathematica Extra Volume ICM 1998

[Mou] F. Moulton, The straight line solutions of the problem of N bodies. Ann. Math. **12**, 1–17 (1910)

[MP] A. Maciejewski, M. Przybylska, Differential Galois approach to the non-integrability of the heavy top problem. Annales de la Faculté des sciences de Toulouse: Mathématiques **14**(1) (2005)

[MPC] H. Müller, A. Peters, St Chu, A precision measurement of the gravitational redshift by the interference of matter waves. Nature **463**, 926–929 (2010)

[MR] J. Marsden, T. Ratiu, *Introduction to Mechanics and Symmetry* (Springer, Berlin, 2003)

[MS] D. McDuff, D. Salamon, *Introduction to Symplectic Topology* (Oxford University Press, Oxford, 1999)

[MV] B. Marx, W. Vogt, *Dynamische Systeme: Theorie und Numerik* (Akad. Verlag, Spektrum, 2010)

[MW] J. Marsden, A. Weinstein, Reduction of symplectic manifolds with symmetry. Rep. Math. Phys. **5**, 121–130 (1974)

[Nar] H. Narnhofer, Another definition for time delay. Phys. Rev. D **22**, 2387–2390 (1980)

[Nat] F. Natterer, *The Mathematics of Computerized Tomography* (Teubner, 1986)

[Ne] I. Newton, Philosophiæ Naturalis Principia Mathematica (1687)

[Ni] L. Nicolaescu, *An Invitation to Morse Theory*, Universitext (Springer, New York, 2007)

[No] E. Noether, *Invariante Variationsprobleme* (Nachrichten von der Königl. Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse, 1918)

[NT] H. Narnhofer, W. Thirring, Canonical scattering transformation in classical mechanics. Phys. Rev. A **23**, 1688–1697 (1981)

[OP] M.A. Olshanetsky, A.M. Perelomov, Classical integrable finite dimensional systems related to Lie algebras. Phys. Rep. **71**, 313–400 (1981)

[OV] A. Onishchik, E. Vinberg, *Lie Groups and Lie Algebras III: Structure of Lie Groups and Lie Algebras*. Encyclopaedia of Mathematical Sciences 41 (Springer, Berlin, 1994)

[Pai] P. Painlevé, Leçons sur la théorie analytique des équations differentielles. Leçons de Stockholm, in Œuvres de P. Painlevé I, Editions du C.N.R.S., Paris (1972), pp. 199–818

[Pat] G. Paternain, *Geodesic Flows*, Progress in Mathematics (Birkhäuser, Basel, 1999)

[PdM] J. Palis, W. de Melo, *Geometric Theory of Dynamical Systems* (Springer, New York, 1982)

[Pen] R. Penrose, The apparent shape of a relativistically moving sphere. Proc. Cambridge Philos. Soc. **55**, 137–139 (1959)

[Per] L. Perko, Differential Equations and Dynamical Systems (Springer, 2. Aufl., Berlin, 1991)

[Poe] J. Pöschel, Integrability of Hamiltonian systems on cantor sets. Commun. Pure Appl. Math. **35**, 635–695 (1982)

[Poi1] H. Poincaré, Sur les courbes définies par les équations différentielles. Journal des Mathématiques pures et appliquées **3**, 375–422 (1881)

[Poi2] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste*, vol. 3 (Gauthier-Villars, Paris, 1899)

[Poi3] H. Poincaré, The present and the future of mathematical physics. Bull. Am. Math. Soc. **12**, 240–260 (1906)

[PR1] R. Penrose, W. Rindler, *Spinors and Space-Time: Volume 1, Two-Spinor Calculus and Relativistic Fields* (Cambridge University Press, Cambridge, 1987)

[PR2] I. Percival, D. Richards, *Introduction to Dynamics* (Cambridge University Press, Cambridge, 1983)

[Ra] P.H. Rabinowitz, Variational methods for Hamiltonian systems, in *Handbook of Dynamical Systems*, vol. 1, Part A, ed. by B. Hasselblatt, A. Katok (2002), pp. 1091–1127

[Ro] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos* (CRC Press, 1999)

[Roe] O. Rømer, Demonstration touchant le mouvement de la lumière trouvé. Journal des Sçavans **7**, 276–279 (1676)

[Qu] E. Quaisser, *Diskrete Geometrie Spektrum* (Akad. Verlag, Heidelberg, 1994)

[Sa] D. Saari, *Collisions, Rings, and Other Newtonian N-Body Problems* (AMS, Providence, 2005)

[SB] H. Schulz-Baldes, Sturm intersection theory for periodic Jacobi matrices and linear Hamiltonian systems. Linear Algebra Appl. **436**, 498–515 (2012)

[Schm] St. Schmitz, Zum inversen Streuproblem der klassischen Mechanik. Dissertation, TU München (2006)

[Scho] M. Schottenloher, *Geometrie und Symmetrie in der Physik: Leitmotiv der mathematischen Physik* (Vieweg, Wiesbaden, 1995)

[Schw] M. Schwarz, *Morse Homology*. Progress in Mathematics 111 (Birkhäuser, Basel, 1993)

[Sib] K.F. Siburg, *The Principle of Least Action in Geometry and Dynamics*. Lecture Notes in Mathematics 844 (Springer, Berlin, 2004)

[Sim] B. Simon, Wave operators for classical particle scattering. Commun. Math. Phys. **23**, 37–48 (1971)

[Sin] Ya.G. Sinai, Dynamical systems with elastic reflections. Russ. Math. Surv. **25**, 137–191 (1970)

[SM] C.L. Siegel, J. Moser, *Lectures on Celestial Mechanics*. Grundlehren der mathematischen Wissenschaften 187 (Springer, Berlin, 1971)

[Smi] U. Smilansky, The classical and quantum theory of chaotic scattering, in *Chaos and Quantum Physics*. Les Houches LII, ed. by M.-J. Giannoni et al. (North-Holland, Amsterdam, 1989)

[Sm1] S. Smale, Differentiable dynamical systems. Bull. Am. Math. Soc. **73**, 747–817 (1967)

[Sm2] S. Smale, Topology and mechanics, Part I. Inventiones Mathematicae **10**, 305–331 (1970); Part II: The planar n-body problem. Inventiones Mathematicae **11**, 45–64 (1970)

[Sob1] D. Sobel, *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time* (Harper Collins, 2005)

[Sob2] D. Sobel, *Galileos Daughter: A Drama of Science, Faith and Love* (Frank R Walker Co, 2011)

[Som] A. Sommerfeld, *Mechanics*. Lectures on Theoretical Physics (Academic Press, 1952)

[Spi] M. Spivak, *Calculus on Manifolds* (Addison-Wesley, Reading, 1965)

[StSc] E. Stiefel, G. Scheifele, *Linear and Regular Celestial Mechanics*. Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen 174 (Springer, Berlin, 1971)

[SV] J. Sanders, F. Verhulst, *Averaging Methods in Nonlinear Dynamical Systems* (Springer, Berlin, 1985)

[SW] H. Sussmann, J. Willems, 300 years of optimal control: from the brachystochrone to the maximum principle. IEEE Control Syst. Mag. **17**, 32–44 (1997)

[Ta] S. Tabachnikov, *Geometry and Billiards* (American Mathematical Society, 2005)

[Te] J. Terrell, Invisibility of the Lorentz contraction. Phys. Rev. **116**, 1041–1045 (1959)

[Th1] W. Thirring, *A Course in Mathematical Physics 1: Classical Dynamical Systems* (Springer, Wien, 1978)

[Th2] W. Thirring, *A Course in Mathematical Physics 2: Classical Field Theory* (Springer, Wien, 1998)

[Un] A. Ungar, The relativistic velocity composition paradox and the Thomas rotation. Found. Phys. **19**, 1385–1396 (1989)

[Wa1] W. Walter, *Ordinary Differential Equations*, Graduate Texts in Mathematics 182 (Springer, New York, 1998)

[Wa2] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics 79 (Springer, Berlin, 1982)

[WDR] H. Waalkens, H.R. Dullin, P. Richter, The problem of two fixed centers: bifurcations, actions, monodromy. Phys. D: Nonlinear Phenom. **196**, 265–310 (2004)

[Weid] J. Weidmann, *Spectral Theory of Ordinary Differential Operators*. Lecture Notes in Mathematics 1258 (Springer, Berlin, 2009)

[Wein] A. Weinstein, Symplectic geometry. Bull. Am. Math. Soc. **5**, 1–13 (1981)

[Wey] H. Weyl, Über die Gleichverteilung von Zahlen mod. eins. Mathematische Annalen **77**, 313–352 (1916)

[Whi] E. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge, 1904)

[Wil] J. Williamson, On the algebraic problem concerning the normal forms of linear dynamical systems. Am. J. Math. **58**, 141–163 (1936)

[Wit] E. Witten, Supersymmetry and Morse theory. J. Differ. Geom. **17**, 661–692 (1982)

[WPM] J. Wisdom, S. Peale, F. Mignard, The chaotic rotation of Hyperion. Icarus **58**, 137–152 (1984)

[Wu] R. Wüst, Mathematik für Physiker und Mathematiker, vol. 1 (Wiley-VCH, 2009)

[Xi] Z. Xia, The existence of noncollision singularities in Newtonian systems. Ann. Math. **135**, 411–468 (1992)

[Yo] H. Yoshida, Construction of higher order symplectic integrators. Phys. Lett. A **150**, 262–268 (1990)

[Zee] E. Zeeman, Causality implies the Lorentz group. J. Math. Phys. **5**, 490–493 (1964)

[Zeh] E. Zehnder, *Lectures on Dynamical Systems*, Textbooks in Mathematics (European Mathematical Society, Zürich, 2010)

[Zei] E. Zeidler, *Oxford Users' Guide to Mathematics* (Oxford University Press, 2004)

[ZP] W. Zurek, J. Paz, Why We Don't Need Quantum Planetary Dynamics: Decoherence and the Correspondence Principle for Chaotic Systems. arXiv.org (1996)

Table of Symbols

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

Image Credits

- The image on page 1 is due to Cambridge University Library.
- The images in Figure 1.1 on page 10 are due to Andy Wolski (Liverpool).
- The photos on page 11 are from Chapter 10 (written by D. Fowler and P. Prusinkiewicz) of the book [FP] by H. Meinhardt.
- Image on page 137: The U.S. National Archives and Records Administration.
- Image on page 155 (parabola slides): Zentrum Mathematik (Technical University of Munich, Germany).
- Figure 8.6.2 on page 183 from Wikipedia, https://es.wikipedia.org/wiki/Archivo:Desert_mirage_62907.JPG, May 2007, photo courtesy of Mila Zinkova.
- The picture on page 191 is due to Rick Hanley.
- Image of Foucault's pendulum on page 241: Miami University (Oxford, Ohio). Photographer: Scott Kissell
- The photo on page 277 (Noé Lecocq) is from Wikimedia. https://commons.wikimedia.org/wiki/File:Billard.JPG, August 2006, photo by Noé Lecocq in collaboration with H. Caps. Courtesy of Noé Lecocq.
- The graphics on page 325 is due to Ulrich Pinkall.
- Picture on pages 365, 383: courtesy of NASA/JPL-Caltech.
- The photo on page 372 (left) was taken by Tunç Tezel.
- The pictures on page 379 are from
  - The Mathematica Journal, Special Section: Dynamic Rotation of Space Station Mir, Oct 1999 (left),
  - NASA/JPL/Space Science Institute (center) and
  - The Mathematical Sciences Research Institute (MSRI, Berkeley, California), DVD 'The Right Spin' (right).
- The image on page 384 is due to Gérard Lacz.
- The picture on page 391 is from NASA/JPL-Caltech.
- The graphics on page 428 is taken from Figure 8.3-3 in the book *Foundations of Mechanics* [AM] by Ralph Abraham and Jerrold E. Marsden.

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

- Left image on page 435: National Archaeological Museum, Athens (Greece) (NAM inv. No. X 15087); The rights of the depicted monuments belong to the Greek Ministry of Culture and Sports (Law 3028/2002).
  Right image: De Solla Price, courtesy of *Transactions of the American Philosophical Society*, Vol. 64, No. 7 (1974)
- The images on page 441 are by Ute Kraus and Marc Borchers, http://www.spacetimetravel.org.
- The image on page 469 is due to Norbert Nacke.
- The image on page 503 is by Oberwolfach Research Institute for Mathematics.
- Image on page 589: courtesy of Wikipedia author RokerHRO.
- The graphics on page 610 is due to Christoph Schumacher.
- The figure on page 640 is due to Markus Stepan.

I am grateful to all who allowed to reprint the mentioned figures.
The remaining figures were produced by the author.

Index of Proper Names

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

Index

A. Knauf, *Mathematical Physics: Classical Mechanics*, La Matematica per il 3+2 109,
https://doi.org/10.1007/978-3-662-55774-7

H

I